



**UNIVERSIDAD DE CHILE  
FACULTAD DE CIENCIAS SOCIALES  
ESCUELA DE POSTGRADO**

**Del *big data* a la Ciencia de Datos: el trabajo investigativo de las Ciencias Sociales latinoamericanas en la era de la digitalización de los datos.**

Tesis para optar al grado de doctor

**BÁRBARA MAGALHÃES DE AGUIAR OLIVEIRA**

Director(a):

**Andrés Gomez Seguel**

Comisión Examinadora:

**Dr. Pablo Kreimer**

**Dr. Bernardo Amigo**

**Dr. Pablo Cottet Soto**

**Santiago de Chile, año 2024**

## **Dedicatoria**

A mis compañeros de doctorado,  
caminantes desbravadores del  
mundo de la investigación social.

## **Agradecimientos**

Estos últimos cinco años he aprendido bastante sobre gratitud. El doctorado me abrió un nuevo mundo. Soy mujer, extranjera, académica, de habla portuguesa, viviendo en Chile, este país un poco gris y frío, ligeramente reacio al forastero, pero al cual debo tanto: mi profesión, mi familia, mi hijo. Soy grata.

El programa de doctorado me abrió puertas. Puertas profesionales que me permitieron, gracias a la vida, como diría Violeta Parra, y a las bonitas coincidencias de esta vida, a estar insertada en uno de los trabajos más interesantes y pioneros que pudiera un estudiante doctoral participar.

Para empezar, agradezco a los recursos públicos, a través de la beca ANID, fundamental para proporcionar dignidad material y financiera durante los estudios, sobre todo en tiempos de protestas y pandemia.

Estoy extremadamente agradecida a mis compañeros, mis valientes colegas de profesión, que igual que yo pasaron por demasiados altos y bajos en esta jornada doctoral. Fue emocionante estar estudiando la sociedad chilena, cada uno a su manera, mientras la propia sociedad se daba vuelta entera.

Me agradezco a mí misma por la voluntad y perseverancia diaria, especialmente por el hecho de emprender dicha investigación mientras cuidaba a mi pequeño hijo. Agradezco a la lactancia por brindarme claridad en mis ideas.

Agradezco a mi compañero de vida, Tonho, por su apoyo incondicional, por ayudarme en todo lo que puede y por acceder a salir de la casa casi todos los fines de semana para brindarme silencio y la quietud necesaria para escribir este trabajo.

A mis suegros, por cuidar a Caetano cuando necesitaba salir a trabajar, y por hacerme familia.

A mi profesor guía, Andrés Gómez, por ser más que un tutor, un colega.

A Henry, informante clave, colaborador y compañero de investigación. Responsable por rescatarme de la matrix y devolveme la potencia de la investigación en contexto de presencialidad.

También agradezco a mi padre por enseñarme a tornar la escritura de este texto más eficiente y a mi madre por su constante apoyo y oraciones.

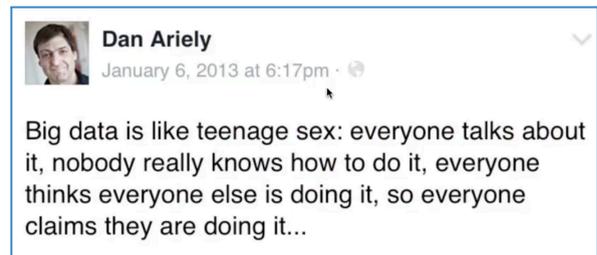
Gracias a ti, lector, por el interés en acompañar esta lectura.

Espero que disfruten. A mi me encantó escribirla.

¡Saludos!



## Epígrafe



El *big data* es como el sexo adolescente:  
todo el mundo habla de ello, nadie sabe realmente  
cómo hacerlo, todo el mundo piensa que  
todos los demás lo están haciendo,  
así que todo el mundo afirma que lo está haciendo.

Dan Ariely, 6 de enero de 2013 a las 6:17 PM en twitter

## Índice

<b>Dedicatoria.....</b>	<b>2</b>
<b>Agradecimientos.....</b>	<b>3</b>
<b>Epígrafe.....</b>	<b>6</b>
<b>Índice.....</b>	<b>7</b>
<b>Resumen.....</b>	<b>13</b>
<b>Capítulo 1: Ruta de la investigación.....</b>	<b>1</b>
1.1 Introducción.....	1
1.2 Contexto del estudio.....	3
1.3 Objetivos y pronósticos.....	7
1.4 Nota a los lectores: el trabajo de campo.....	12
<b>Capítulo 2: Discusión teórica: lentes para el campo.....</b>	<b>20</b>
2.1 El fenómeno big data.....	20
2.2 El (big) dato digital.....	23
2.3 Big data y las investigaciones en ciencias sociales.....	25
2.4 La investigación científica social.....	31
2.5 La investigación social latinoamericana.....	36
<b>Cap 3: Scielo, una librería científica latinoamericana.....</b>	<b>43</b>
3.1 Parámetros de búsqueda: como se hizo para obtener tales resultados.....	45
3.2 Resultados generales.....	55
3.2.1 Extensión bibliométrica.....	55
3.2.2 Tasa de crecimiento.....	56
3.2.3 Textos metodológicos.....	57
3.2.4 Textos Teóricos.....	59
3.2.5 Autores.....	61
3.3.6 Género.....	63
3.2.7 Colaboración:.....	64
3.2.8 Lista de textos más relevantes (más descargados).....	65
3.2.9 Idiomas:.....	68
3.2.10 Áreas disciplinares de la revistas:.....	70
<b>Cap 4: Scopus: ampliando el escopo de investigación.....</b>	<b>72</b>
4.1 Scopus: la base de datos más grande en latinoamérica.....	75
4.2 Parámetros e índices para la aplicación de una metodología bibliométrica compartida.....	76
4.3 Resultados encontrados.....	79
4.3.1 Big data en los distintos continentes.....	84
4.3.2 Autores que publican temáticas big data en las ciencias sociales.....	89
4.3.3 Instituciones más relevantes para la difusión de temas big data en las Ciencias Sociales.	93
4.3.4 Revistas y conferencias relevantes sobre big data en las Ciencia Sociales.....	95
4.3.5 Citas más relevantes.....	103
4.3.6 Tendencias de las palabras clave a lo largo del tiempo en las ciencia sociales mundiales..	107

4.3.7	Palabras claves en los textos Latinoamericanos en las ciencias sociales.....	109
4.3.8	Redes de colaboración.....	113
4.3.9	Conclusiones observadas a partir del estudio bibliométrico en Scopus.....	117
<b>Cap 5:</b>	<b>Scielo y Scopus: que los textos nos tiene a decir.....</b>	<b>119</b>
5.1	El procesamiento computacional de textos como el origen de la manipulación de macrodatos sociales.....	122
5.2	La historia: reinterpretando el pasado a través de técnicas big data.....	124
5.3	Economía y negocios: una interpretación macro social de la vida económica a través de big data.....	125
5.4	La psicología y la intervención en salud a partir de big data: recalibrando la disciplina.	127
5.5	Las ciencias de la educación y los nuevos procesos de enseñanza a partir de big data....	128
5.6	La sociología y su discurso crítico cuanto a la introducción de técnicas big data en el desarrollo del saber social.....	129
5.7	El derecho y el enfoque jurídico garantizando la legalidad de los procesos investigativos a través de big data.....	131
5.8	Ciencia política, opinión pública y participación ciudadana impulsadas por herramientas big data.....	132
5.9	Geografía social y económica: remodelando los espacios geográficos a través de big data..	133
5.10	Periodismo y comunicación: las implicaciones de big data frente a la información digital.	134
<b>Cap 6:</b>	<b>Ruta etnográfica: contexto de investigación.....</b>	<b>139</b>
6.1	La etnografía sensitiva y digital como enfoque etnográfico para los estudios de laboratorio involucrados en la temática big data.....	143
6.2	El laboratorio como herramienta de análisis social.....	151
6.3	GobLab: un laboratorio social en big data y ciencias de datos.....	155
6.3.1	Uso de datos y la innovación pública: como se construye un laboratorio de ciencia de datos de gobierno.....	162
6.3.2	Formar, incidir y investigar: la tríada que sostiene la misión del laboratorio.....	163
6.3.3	Colaboradores diversificados como fortaleza institucional.....	165
6.3.4	La formación y capacitación humana como práctica de consolidación de un saber.....	169
6.3.5	La Investigación aplicada como principio guía de las prácticas del laboratorio.....	171
6.3.6	La incidencia como mecanismo de aporte social.....	174
6.3.7	Publicaciones relevantes ni siempre son artículos: el texto y sus múltiples funcionalidades.....	176
6.3.8	Internacionalización como prerrogativa de existencia: marcas de una práctica mundializada.....	177
<b>Cap 7:</b>	<b>La ciencia de datos es el nuevo big data.....</b>	<b>180</b>
7.1	Big data versus ciencia de datos? diferenciando conceptos aparentemente similares.....	191
7.2	El big data como una técnica presente en la ciencia de datos.....	195
<b>Capítulo 8:</b>	<b>Investigaciones en un laboratorio en ciencias de datos.....</b>	<b>200</b>
8.1	Esfera de datos (E1): los datos como centro de las investigaciones aplicadas.....	203
8.2	Esfera Problema (E2): la resolución de problemas como objetivo principal de los proyectos.....	210
8.2	Espacio-Tema (ETem 6): unificando los diversos intereses, saberes y soluciones.....	213
8.3	Espacio-teoría (ETeo10): los enmarcado teóricos como mecanismo de validación de los resultados.....	214

8.3 Profesionales: los humanos por detrás del funcionamiento de la máquina laboratorial.....	220
8.4 Un sistema socio-técnico: la manera GobLab de funcionar.....	222
8.5 Qué puede enseñarnos un laboratorio social en ciencia de datos: Implicaciones metodológicas a la investigación social con vocación data science.....	229
8.5.1 Focalización en la investigación aplicada.....	230
8.5.2 Diseño de investigación limitada a la disponibilidad de los datos.....	233
8.5.3 Colaboración e interdisciplinariedad.....	238
8.5.4 Tolerancia y confianza en la imprevisibilidad.....	244
8.5.5 La representatividad de la muestra pierde relevancia.....	253
8.5.6 Cuantificación y visualización.....	262
<b>Cap 9: (Posibles) Implicaciones sociales de una sociedad computacional.....</b>	<b>267</b>
9.1 La sociedad computacional: representación de lo social en los estudios de ciencias sociales.....	274
9.1.1 Conectividad.....	278
9.1.2 Tecno-encanto.....	280
9.1.3 Automatización de la vida: eficiencia y aceleración de los procesos.....	287
9.1.4 Velocidad.....	292
9.1.5 Identidad global.....	295
9.1.6 Alienación del trabajo humano por detrás de las máquinas.....	299
9.1.7 Identidad Anfibia.....	303
<b>Cap 10: El científico social y el futuro de la profesión: caminos para la consolidación de un saber en emergencia.....</b>	<b>315</b>
10.1 El profesional de hoy: cómo un ambiente de investigación y artefactos sociotécnicos generan una nueva demanda de profesionales.....	317
10.2 El profesional del mañana y la consolidación de las ciencias sociales computacionales.....	325
10.2.1 Los especialistas: expertos interdisciplinarios integrados en atribuciones emergentes.....	329
10.2.2 Formación: la importancia de multiplicar un campo emergente.....	331
10.2.3 Los Analistas: el ejercicio de la “función humana” en contextos tecno-encantados..	332
10.3 La Creación académica de las Humanidades Digitales y el saber computacional.....	336
10.3.1 La computacionalización de los saberes sociales: ¿Qué son las Humanidades Digitales?	338
10.3.2 Los curriculum: la estructuración académica de las humanidades digitales.....	341
10.3.3 Áreas de aplicación: usando las tecnologías computacionales para investigar la sociedad digital.....	350
10.3.4 Las humanidades digitales en América Latina.....	353
<b>Cap 11: Por unas ciencias sociales computacionales: conclusiones, aprendizajes y límites de la investigación.....</b>	<b>356</b>
<b>Referencias.....</b>	<b>363</b>
<b>Anexos.....</b>	<b>373</b>
<b>I. Glosario.....</b>	<b>373</b>
<b>II. Tablas de referencia por disciplina.....</b>	<b>377</b>

## Índice de tablas y figuras

Figura 1. Tipos y tamaños big data.....	17
Figura 2. Big Data Science versus Big Data Social Science, de 2004 a 2023.....	23
Figura 3. Términos de la búsqueda repentina en 2014.....	24
Figura 4. Portada e índice de Abrir las Ciencias Sociales, 1995.....	37
Figura 5. Mapa general de la investigación.....	62
Figura 6. Definición de las categorías de análisis.....	66
Figura 7. Publicaciones por país de Scielo entre 2013 y 2023.....	71
Figura 8. Scielo Analytics: Proporción x Área temática.....	72
Tabla 1. Comparación entre búsqueda de “big data” en todas las áreas temáticas vs. ciencias humanas.....	74
Tabla 2. Publicaciones sobre big data por país en Scielo.....	74
Figura 9. Base Scielo: “big data” + todas las áreas temáticas.....	75
Figura 10. Base Scielo: búsqueda de “big data” + “ciencias humanas”.....	75
Figura 11. Datos bibliométricos “CS” x “todas las áreas”, hasta 2023.....	76
Figura 12. Nube de las palabras claves de las publicaciones de CS en Scielo.....	79
Figura 13. Categorías/variables de la base de datos oficial de Scielo.....	82
Figura 14. Atributos primer texto en CS y big data, 2014.....	86
Figura 15. Publicaciones en BD por cada año + tasa de crecimiento.....	87
Tabla 3. Total de términos BD por tipo de texto.....	88
Figura 16. Datos generales: textos de tipo metodológico.....	88
Figura 17. Datos generales: textos de tipo metodológico.....	89
Tabla 4. Cantidad de autores por país de publicación.....	91
Tabla 5. Comparación publicaciones entre hombres y mujeres.....	93
Tabla 6. Colaboración entre autores.....	94
Figura 18. Listado de textos más descargados en la plataforma.....	95
Figura 19. Principales términos asociados a las publicaciones más relevantes.....	95
Figura 20. Proporción de palabras claves en Scielo.....	96
Tabla 7. Frecuencia de términos más comunes de la base de datos.....	97
Figura 21. Índices de las publicaciones en español.....	98
Figura 22. Índices de las publicaciones en portugués.....	99
Figura 23. Congreso ESOCITE- LALICS, mesa 31.....	103
Figura 24. Número total de publicaciones en ciencias sociales sobre big data por año, 1970-2019..	111
Figura 25. Publicaciones sobre big data por país con mayor número de publicaciones y año, 1970-2019.....	114
Figura 26. Publicaciones sobre big data por año y por región.....	116
Figura 27. Publicaciones sobre big data en ciencias sociales en los países latinoamericano.....	117
Figura 28. Publicaciones sobre big data por país y grupo de ingresos.....	118
Tabla 8. Los 10 principales autores de las publicaciones en ciencias sociales sobre big data en el mundo.....	121
Tabla 9. Los 5 principales contribuyentes por nivel de renta en el país.....	121
Tabla 10. Principales autores por país que investigan sobre big data en América Latina.....	122

Tabla 11. Las 5 principales instituciones que investigan sobre big data por nivel de ingresos y país	124
Tabla 12. Las 5 principales instituciones que investigan sobre big data en América Latina.....	125
Figura 29. Principales actas de congresos y revistas científicas que publican investigaciones sobre big data en ciencias sociales (a + b).....	126
Tabla 13. Principales revistas científicas que publican artículos relacionados con big data en ciencias sociales por grupo de países.....	128
Tabla 14. Principales revistas científicas que publican artículos relacionados con big data en ciencias sociales por región.....	129
Tabla 15. Principales conferencias científicas en las que se han presentado trabajos relacionados con los big data en ciencias sociales por grupo de países.....	130
Tabla 16. Principales conferencias científicas en las que se han presentado trabajos relacionados con los big data en ciencias sociales por regiones.....	131
Tabla 17. Los 10 artículos más citados con la palabra clave “big data” en ciencias sociales.....	133
Tabla 18. Los 10 autores más citados en ciencias sociales sobre big data.....	135
Tabla 20. Los 5 autores de ciencias sociales más citados sobre big data en América Latina.....	136
Figura 30. Evolución de las palabras clave en las publicaciones de ciencias sociales sobre big data de 2010, 2015 y 2019 en todo el mundo.....	138
Figura 31. Palabras clave en las publicaciones de ciencias sociales sobre big data por región....	139
Figura 32. Palabras clave en las publicaciones de ciencias sociales sobre big data por país.....	141
Figura 33. Red de colaboración big data en las ciencias sociales en 2014.....	145
Figura 34. Redes de colaboración internacional por autores.....	146
Figura 35. Redes de colaboración internacional por instituciones.....	147
Figura 36. Descripción y colaboraciones proyecto SAMSARA.....	168
Figura 37. Presentación GobLab, Escuela de Gobierno.....	171
Figura 38. Redes internacionales asociadas a GobLab.....	172
Tabla 21. Objetivos estratégicos GobLab, 2023.....	175
Figura 39. Equipo GobLab, 2022.....	176
Figura 40. Investigadores asociados a UAI en GobLab en 2022.....	177
Figura 41. Colaboradores UAI en GobLab en 2022.....	177
Figura 42. Colaboradores GobLab en 2022.....	179
Figura 43. Formación GobLab UAI - 2022.....	181
Figura 44. Ejemplos de investigación aplicada en GobLab.....	183
Figura 45. Investigaciones GobLab UAI en 2022.....	184
Figura 46. Palabras más frecuentes- Publicaciones GobLab.....	189
Tabla 22. Relación de entrevistas realizadas y analizadas para la investigación.....	194
Figura 47. Evolución comparativa del número de búsquedas de big data y data science.....	198
Figura 48. Búsqueda de los términos big data y data science pr región, desde 2004 a 2023.....	199
Tabla 23. Proporción de búsquedas relacionadas a big data y data science por país.....	201
Figura 49. Relación entre los términos “big data” y “ciencia de datos” en Brasil, 2004 a 2023..	201
Tabla 24. Comparación entre big data y ciencia de datos.....	207
Figura 50. Las siete actividades de las ciencias de datos según H.E Brady (2019).....	209
Figura 51. Email invitación participación BigSurvey.....	212
Figura 52. Laboratorio de ciencia de datos y las investigaciones aplicadas de ciencias sociales.	218
Figura 53. Infraestructura de Operación Sistema Bem-te-vi, 2023.....	220

Figura 54. Línea del tiempo de la capacidad de gestión y gobernanza de datos.....	221
Tabla 25. Capacidad infra computacional versus capacidad de gestión y gobernanza de datos...	223
Tabla 26. Actividades profesionales involucradas en el sistema Investigación social en Ciencia de Datos.....	235
Figura 55. Desafíos IS, Proyecto Fondef SAMSARA, 2022.....	240
Tabla 27. Diferencias entre el modelo clásico y el modelo basado en datos.....	252
Tabla 28. Dimensiones potencializadoras de innovación en proyectos de investigación en ciencia de datos.....	264
Tabla 29. El papel del error en proyectos en ciencias de datos en el marco de las ciencias sociales..	267
Tabla 30. Potencialidades y desafíos relacionados a la pérdida de relevancia de la representatividad de la muestra.....	275
Tabla 31. Paradigmas de la ciencia de datos y la sociedad computacional.....	286
Tabla 32. Resumen esquemático conectividad.....	291
Tabla 33. Resumen esquemático tecno-encanto.....	299
Tabla 34. Resumen esquemático automatización de la vida.....	303
Tabla 35. Resumen esquemático de velocidad.....	306
Tabla 36. Resumen esquemático identidad global.....	310
Tabla 37. Resumen alienación del trabajo humano.....	314
Tabla 38. Resumen esquemático identidad anfibia.....	326
Tabla 39. Principales rasgos del nuevo profesional en ciencias sociales.....	330
Tabla 40. Listado de actividades del nuevo científico social computacional.....	339
Tabla 41. Listado de disciplinas optativas en posgrados en Humanidades Digitales.....	358
Figura 56. Nube de palabras especialidad cuerpo docente.....	360
Tabla 42. Aplicabilidades de los recursos computacionales en diferentes disciplinas de las humanidades digitales.....	361
Tabla 43. Listado de organizaciones en humanidades digitales.....	364

## Resumen

Esta investigación aborda la introducción del término "big data" en las ciencias sociales latinoamericanas, evaluando su impacto y las implicaciones sociales de la adopción de tecnologías computacionales. Se realiza una exhaustiva revisión en bases de datos como Scopus y Scielo, seguida de una investigación etnográfica basada en entrevistas y observación participante.

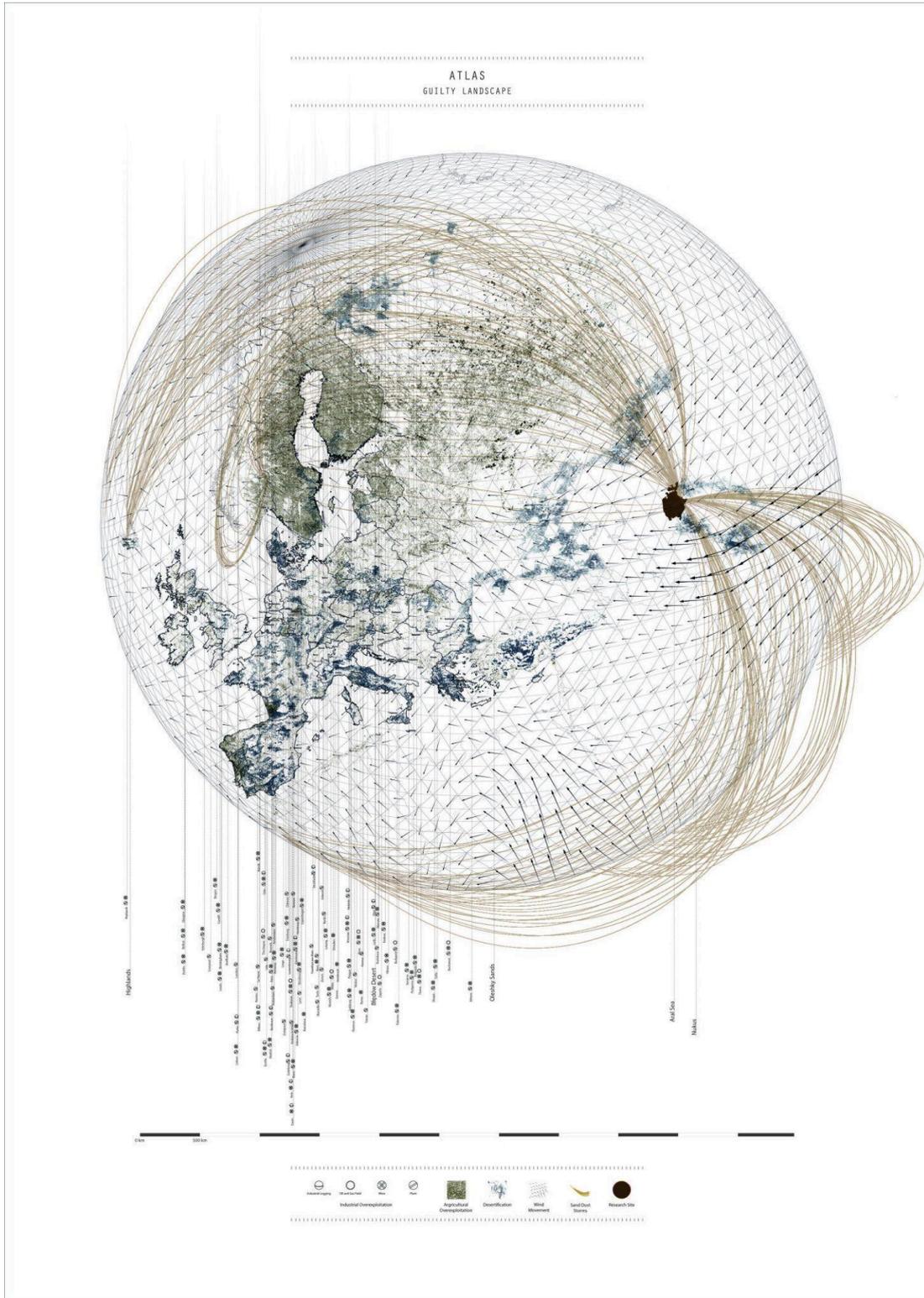
El análisis bibliométrico revela una escasez de publicaciones en ciencias sociales sobre big data en Latinoamérica, las cuales son principalmente teóricas y centradas en discutir aplicaciones y limitaciones. La investigación etnográfica muestra que la adopción de big data se asocia con términos científicos y debates macro políticos.

Se destaca la evolución del término "big data" a "ciencia de datos" y su consolidación en investigaciones sociales. Se evidencia una falta de publicaciones con resultados de investigaciones y un enfoque teórico predominante. La investigación etnográfica en un laboratorio de innovación pública resalta cambios en prácticas científicas y la emergencia de las humanidades digitales.

Se observa la formación de una sociedad computacional, afectando la estructura social en ámbitos como economía, cultura y trabajo. Se identifica un nuevo profesional en ciencias sociales adaptado a las transformaciones y surge el campo disciplinario de las humanidades digitales.

Este estudio contribuye al entendimiento de las intersecciones entre innovación tecnológica y dinámicas sociales en América Latina, ofreciendo una visión detallada de las transformaciones en la investigación social en la sociedad computacional. Su relevancia para los estudios CTS latinoamericanos reside en promover un diálogo crítico y enriquecedor sobre la interacción entre ciencia, tecnología y sociedad en la era digital

# PARTE 1: REPORTE



## Capítulo 1: Ruta de la investigación

### 1.1 Introducción

Era el año 2019, más precisamente octubre, Santiago ardía en llamas y nosotros necesitábamos estudiar. Por estas cosas que pasan en la vida, 2019 es el mismo año que inicio mi jornada doctoral también con un hijo de 4 meses en mis brazos.

Estudiaba en Juan Gomez Millas, conocido por su poder de movilización, lucha y resistencia. Las tomas ya eran una constante antes del estallido social, después de octubre, ya no era viable ni posible, volver físicamente ahí. Así que fuimos invitados a la digitalidad. Clases, textos y referencias, reuniones y trabajos de fin de curso, todo, en una pantalla.

Mi emersión en la problemática de la digitalidad y la ciencia, comenzó un poco antes de lo que empezó para el resto del globo, en marzo de 2020. Sí, estamos hablando de Covid-19 instalándose en nuestro modo de vivir. ¡Ni siquiera había partido mi 3er semestre!

Las provocaciones y cuestionamientos que dan origen a este trabajo nacen de la experiencia de una digna revuelta popular, *chacoalhando* (permítame el uso del portugués, mi idioma madre) mentes, deseos y esperanzas, sobre todo para mejor.

Un día, mientras mi marido estaba presencialmente en las protestas, yo, en casa con mi hijo pequeñito en los brazos, decido prender la computadora para ver la tele: mientras se reportaba las marchas y transmitía en vivo la manifestación, algunos invitados comentaban su parecer sobre los acontecimientos en todo el país. Le tocó la palabra a una persona, que al explicar los diversos motivos que según ella fueron la gota de agua para la revolución, fue el alza del precio del metro, tras años consecutivos.

Mientras explicaba, la persona en cuestión, dice algo como: “Cuando interrogamos a las autoridades competentes en base a qué se calculaban los precios del pasaje, los costos, el poder de compra de la gente, para por fin decidir cuánto más si podría subir el precio del metro, nos contestó diciendo que ellos trabajaban con un equipo de ingenieros que junto con los algoritmos desarrollados, llegaban a la decisión del costo final de los pasajes.”

Muy enrabiada, seguía su raciocinio: “no es posible que una decisión de esta magnitud, del precio de una cuestión que afecta toda a la gente, de la ciudad más grande del país, de unas de las flotas de metro más grandes de América Latina, sea tomada por un grupito de ingenieros metidos a gestores públicos y computines. ¡Y aún decir que los algoritmos decidieron por el

valor y mejor momento del alza! Esto está demás. A mi me gustaría conversar con estos algoritmos y preguntarles, por qué creen que pueden año tras año aumentar los costos de vida de la población sin mirar a la gente”.

Este discurso me marcó inmensamente. No sólo porque dicha persona tenía un poco de razón, también porque fue la primera vez que me di cuenta, de verdad, de verdad, que sí: nuestras vidas colectivas, como ciudadanos de un gobierno, están impactados y sometidas a toma de decisiones algorítmicas, o automatizadas, como algunos prefieren decir. Si es así con el metro, también es así con la salud, la vivienda, los impuestos, concluí.

Creo que esta entrevista me impactó también porque convivo con un ingeniero programador y también él de alguna forma es el hombre por detrás de una máquina que conjuntamente, sin ni darse cuenta, toma decisiones sobre nuestras vidas. Él no trabaja modelando algoritmos directamente para el gobierno, pero la empresa donde trabaja, nacida de un proyecto de iniciación científica en una de las universidades privadas más importantes del país, influye en la forma de hacer minería en Chile, un país dependiente de esta fuente de ingresos. ¿Estaría él, indirectamente colaborando de igual manera para los ingresos del país, ayudando, sin querer, a establecer el precio de las producciones mineras de Chile, por ejemplo?

Conocía su forma de trabajar, estaba más o menos familiarizada con las tareas y procedimientos de su grupo de trabajo. Convivo también con sus amigos ingenieros matemáticos y programadores, trabajando en distintas áreas de influencia (transporte, banco, salud, agronegocio, minería), cada uno de ellos generando informaciones, grandes cantidades de datos y modelos que impactan directamente la vida de cualquiera de nosotros sin ninguna, digamos, reflexión respecto de los impactos sociales de la generación de información que ellos modelan. ¿Estarían ellos jugando el deporte equivocado, en nuestra consolidada y tradicional cancha de fútbol?

Lo que dijo esa señora fue básicamente que los ingenieros están jugando con la vida de la gente a través de sus algoritmos como si fuesen verdaderos gestores. Estaríamos, nosotros, los científicos sociales, perdiendo campo, espacio y agencia, por un par de algoritmos? Y sé sí, pero, ¿qué estamos haciendo al respecto? Esa misma noche escribí un texto, una crónica con el título: ¡Precisamos conversar con los algoritmos!

Desde ese momento me interesé por investigar sobre la forma en que estos recursos computacionales toman cuenta de nuestras vidas. Mis inquietudes acerca de este tema fueron siendo refinadas y las dudas fueron tomando forma de preguntas de investigación. Algunas lecturas y muchas *googleadas* después, acá estoy, defendiendo mi derecho a ser doctora, más madura en relación a esta investigación, pero con las mismas inquietudes de esta chica de hace 4 años atrás: el lugar de las ciencias sociales, o mejor dicho, de las investigaciones en ciencias sociales en el universo de los grandes datos, de los algoritmos, la computación y la digitalidad.

Considero que mi problema de investigación: “el lugar de big data en las investigaciones de ciencias sociales en latinoamérica”, es fruto de mi entorno, mi experiencia y del deseo aventurarme en las posibilidades que la digitalización de la vida, post pandemia, post estallido social, nos llevó a todos.

## **1.2 Contexto del estudio**

La digitalización de mi mundo y de mi quehacer en cuanto cientista en formación, me llevó a indagar sobre mi propia realidad, sobre mi práctica científica y el futuro de mi profesión. No podemos olvidar que también soy una mujer migrante. Establecerse profesional y financieramente es fundamental para el mantenimiento de la salud mental y dignidad interna, por lo que mi futuro laboral en Chile siempre fue una de mis máximas preocupaciones. Si hay un espacio, un campo de trabajo todavía poco explorado por las ciencias sociales, y que apuntaba para un futuro prometedor y un tanto cuanto inevitable, ¡yo lo quería conocer!

Con un poco de pesquisa en las redes, a finales del año de 2019, no tardé en percibir que la correlación entre el término “algoritmo” y “big data”, estaban, en ese tiempo, separadas por 3 o 4 clics. Hasta ese entonces, no había una clara vinculación entre la investigación científica humanista y el término algoritmo. Sin embargo, la combinación “big data” + “ciencias sociales” solía aparecer al final de estas búsquedas iniciales. Fue así que descubrí las humanidades digitales, o las ciencias sociales digitales, y mi investigación pasó a pertenecer a un nuevo marco conceptual y disciplinario.

Mi sujeto de investigación es, por tanto, las “investigaciones sociales en/con/a través de big data”. Mi motivación y fin es entender las características del trabajo investigativo por detrás

de las investigaciones que utilizan macrodatos para hablar de, comprender, y estudiar aspectos sociales.

Con un universo de posibilidades relativamente grande, aboqué esta investigación a Latinoamérica. No solamente para adaptar las ambiciones al tamaño de una tesis doctoral, sino también porque pude observar, en mi primer acercamiento a la problemática, una escasez de textos académicos escritos en español o en portugués.

Hay, en general, falta de material. Casi todo lo que encontré en ese momento, viene de autores, revistas o congresos del norte global. Mientras buscaba materiales, artículos y textos que trataban de big data en las ciencias sociales, me apunté en un congreso latinoamericano de CTS, ESOCITE (Asociación latinoamericana de Estudios Sociales de la Ciencia y la Tecnología), con el fin de ahondar más. En una mesa de discusión sobre big data en la academia, escuché la presentación de Henry Chávez, un investigador de las ciencias sociales, latinoamericano de Ecuador, postdoctorado en Francia. Ambos estábamos interesados en comprender lo académico por detrás del término big data. Lo contacté y desde entonces hemos estado colaborando. Estudiamos la base de datos de Scopus y Scielo, en busca de publicaciones científicas en ciencias sociales que trabajan, investigan o estudian temas relacionados a los macrodatos en América latina.

Poquísimos días después de procesar los primeros datos de nuestro estudio bibliométrico, me encuentro con un puesto de trabajo ideal para mí: un cargo de “etnógrafo de ciencias de datos” de un proyecto académico-privado, con fuentes de financiamiento de una organización internacional, por supuesto. Nombre del proyecto: Algoritmos éticos, responsables y transparentes. Pasé, entonces, a ser responsable del diseño y conducción de un estudio etnográfico con la finalidad de entender los desafíos de la implementación de sistemas automatizados en diferentes servicios del Estado chileno. ¡De pronto yo trabajaba con aquello que investigaba dentro del área que me proponía entender y estudiar! Es también por esto que a lo largo del tiempo, este trabajo fue ganando la forma de un estudio de naturaleza etnográfica, pues yo estaba naturalmente en terreno. Había que aprovechar de manera estratégica esta afortunada ubicación.

Considero que mi involucramiento y cercanía profesional al tema que me propuse investigar es un plus en la percepción y vivencia del campo. Por ello, quiero insistir y justificar la necesidad de una escritura en primera persona, que según mi perspectiva, no quita objetividad ni científicidad al texto doctoral, sino que evidencia un profundo involucramiento con el

campo de investigación estudiado. Así lo señala mi vínculo con una ciencia social que estudia la Ciencia (con C mayúscula) sin los fantasmas de la prerrogativa de la mirada vertical, desde arriba, imparcial y neutral, del científico observador.

Estar tan inmersa en el núcleo central de mi investigación como trabajadora y científica, aporta una mirada de quien llegando de afuera, logró de verdad estar metida dentro de su universo de investigación. Considero, por lo tanto, tener una perspectiva privilegiada, fresca y dinámica de mi objeto de estudio. Es como me dijo una de las profesoras que nos acompañó de cerca durante toda la jornada doctoral de mi generación: “la vida fue buena contigo porque te está permitiendo investigar una ciencia, un campo profesional, naciente de manera directa y en vivo. Tu investigación se desarrolla al paso que tu campo de actuación nace, se reinventa y se consolida. Esto es estupendo en términos de un trabajo como el tuyo. Estás estratégicamente situada y hay que aprovechar esto” (Marisol Facuse Muñoz, Comunicación personal, mayo de 2022).

Reafirmo lo anterior gracias a lo aprendido de mi guía teórica Karina Knorr-Cetina. La autora es una académica cuyo texto *La fabricación del conocimiento: un ensayo sobre el carácter constructivista y contextual de la ciencia* (2005) problematiza el lugar de neutralidad y objetividad del científico, cuyo trabajo he decidido implementar como referencia bibliográfica y guía de estilo, a modo de reflejar coherencia con lo propuesto.

Defiendo y ejecuto, así, una investigación científica construida en base a los presupuestos teóricos de la cibernética de segundo orden, y emprendo una *etnografía sentiva* (Knorr-Cetina, 2005, p. 91), donde el dejar hablar y dejarse guiar por las casualidades del campo y de las relaciones de investigación, configuran el tono de mi diálogo con el terreno y también con este texto, sin dejar de cumplir con las exigencias de un trabajo de carácter científico.

En esta investigación etnográfica, los hallazgos provienen de la recolección de una gran variedad de formas de datos y archivos, como se supone en los estudios de etnográficos de contexto digital Hine (2015). Aparte del estudio bibliométrico de las bases de datos de Scopus y Scielo, sumé bases más amigables a los latinoamericanos, además de entrevistas con una serie de autores de los artículos encontrados. Participé de congresos, encuentros y seminarios de manera virtual, presencial y grabada. Recolecté material digital en una variedad de formatos y archivos, audios podcast, grabaciones de conferencias, seminarios internacionales disponibles en línea, canales de Youtube de laboratorios y núcleos de

investigación, notas de X<sup>1</sup>, posts de LinkedIn, pantallazos de presentaciones y conferencias, *scraping* de páginas web y grabaciones de grabaciones. Todo esto se entiende como material y, por lo tanto, evidencia, en una investigación en ambiente digital como esta. El análisis de la información aconteció de manera diversificada y muchas veces intuitiva.

Del estudio bibliométrico, resultó una actualización del panorama general de la producción científica de las ciencias sociales. El análisis del texto como documento científico, sirvió para profundizar en la comprensión de la narrativa de la ciencia. Por otro lado, el análisis de contenido a partir de la transcripción de las entrevistas y las nubes de palabras fueron estrategias utilizadas, por ejemplo, para materiales como grabaciones de entrevista, conferencias y materiales audiovisuales.

Las imágenes, gráficas, capturas de pantalla, páginas web o materiales gráficos, fueron categorizadas en una planilla personal, según su lugar de procedencia, tema, tipo de evidencia, a lo largo de la construcción del relato etnográfico para luego encajar en cada temática o categoría de análisis más o menos parecido a la confección de un gran mosaico o rompecabezas. El campo habla primero, su interpretación y categorización se va ajustando conforme las evidencias van siendo organizadas y mi pensamiento analítico, así, produce.

En medio de tanta diversidad de información y formas distintas de categorización y análisis, para garantizar objetividad en la aprehensión de los contenidos, me apoyé en la observación sistemática de 5 categorías o dimensiones principales: 1- Investigación científica en ciencias Sociales; 2- Trabajo Científico; 4- big data/algoritmos; 5- América Latina, y finalmente, 6- ciencia de datos, como explicaré más adelante.

Los resultados de este trabajo son parte de una trayectoria de investigación de la que estoy muy orgullosa y agradecida. Estar vinculada a este tema me abrió muchas puertas profesionales que, a la vez, me abrieron más caminos, contenidos y experimentaciones, que aportaron de manera única a la investigación de este trabajo doctoral. Y es justamente sobre esto que quiero contar en el texto subsiguiente: Pasados 4 años de investigación, ¿qué les puedo contar sobre lo que descubrí? Como diría Andrés, mi querido y paciente profe-guía: “¿qué pasa con el big data en las ciencias sociales de latinoamérica, Bárbara?!”

---

<sup>1</sup> X acá se refiere al antiguo Twitter.

### 1.3 Objetivos y pronósticos

Los objetivos propuestos para esta investigación fueron diseñados en base a la voluntad por conocer lo que se estaba produciendo en la literatura académica sobre big data, específicamente, en las investigaciones de ciencias sociales en latinoamérica. Por ello, el sujeto de este estudio son las investigaciones en ciencias sociales<sup>2</sup>.

#### Objetivos propuestos en el proyecto de cualificación

Conocer las representaciones sobre lo social respecto al uso de big data en las publicaciones científicas de ciencias sociales latinoamericanas.

#### Objetivos específicos

- a) Mapear la producción latinoamericana en ciencias sociales que utilizan big data en su investigación, a través del análisis bibliométrico de las publicaciones científicas entre los años 2009 a 2020, en contraste a la producción mundial;
- b) Identificar los términos, conceptos y contenidos en evidencia y asociados al uso de big data en las publicaciones científicas de las ciencias sociales latinoamericanas durante este mismo período;
- c) Analizar las representaciones (interpretaciones) de lo social contenidas en las publicaciones científicas en Latinoamérica y que discuten la problemática big data en sus investigaciones, entre los años 2009 y 2020.

Este estudio se inicia a partir del entendimiento que la emergencia de lo digital en la vida cotidiana favoreció la masificación de la producción de datos sobre los más variados sectores de la vida social. Así como el mundo científico se apropia de estos datos para investigar, modelar y pronosticar comportamientos de agencia no humanos, también las ciencias sociales comienzan a utilizar los grandes datos para investigar y teorizar sobre lo social, sobre los humanos. Como consecuencia, la utilización de big data implica cambios en el comportamiento investigativo en ciencias sociales y en el modo en que se concibe lo social (Nascimento, 2016). Una vez que entendemos que la ciencia y la producción del

---

<sup>2</sup> Se considerará disciplinas pertenecientes a las ciencias sociales los trabajos provenientes de las siguientes disciplinas según la OCDE: i) psicología; ii) economía y negocios; iii) ciencias de la educación; iv) sociología; v) derecho; vi) ciencias políticas, vii) geografía social y económica y viii) periodismo y comunicación.

conocimiento científico es construido como una vía de doble afectación. La sociedad transforma sus instrumentos y tecnologías, y con ello la ciencia y lo que se sabe científicamente respecto a la sociedad también se actualiza.

Por medio del mapeo de la producción científica se espera encontrar un campo interdisciplinar tanto respecto a los colaboradores (científicos y laboratorios), como a las actividades y contextos metodológicos de la práctica investigativa. De igual manera, se estima que las investigaciones producidas bajo el marco big data, estarán fuertemente vinculadas al paradigma computacional incorporando el *modus operandi* investigativo de las ciencias exactas, donde la valorización de las estrategias de análisis de datos basadas en texto estará evidenciada.

Como ya señalan las teorías decoloniales (Mignolo, 2008; Quijano, 2014; Domingues, 2015), se espera la construcción de un pensamiento social latinoamericano aun estrechamente vinculado a la metrópoli y los grandes centros de investigación de los países del Norte global. Respecto al contexto de producción de las investigaciones sociales en big data, se presupone que las publicaciones analizadas serán mayoritariamente masculinas, angloparlantes y provenientes de la relación con centros de investigación de países del norte.

Por último, la apuesta con respecto a las representaciones de lo social concebidas a través del uso de big data es que su prerrogativa sería la búsqueda de las problemáticas globales y un mayoritario interés por los análisis de fenómenos complejos.

Este estudio adopta un enfoque exploratorio y reflexivo, dado el carácter experimental del campo investigado: las ciencias sociales computacionales, aún en consolidación. El objetivo, por tanto, es más demostrativo y exploratorio que conclusivo.

Se aplicaron metodologías mixtas, divididas en dos fases. La Fase 1 consistió en un estudio bibliométrico de textos académicos latinoamericanos en ciencias sociales, recopilados en las bases Scielo (2014-2020) y Scopus (2009-2021). Este estudio exploró fenómenos del campo sin pretender predecir ni demostrar conclusiones rígidas.

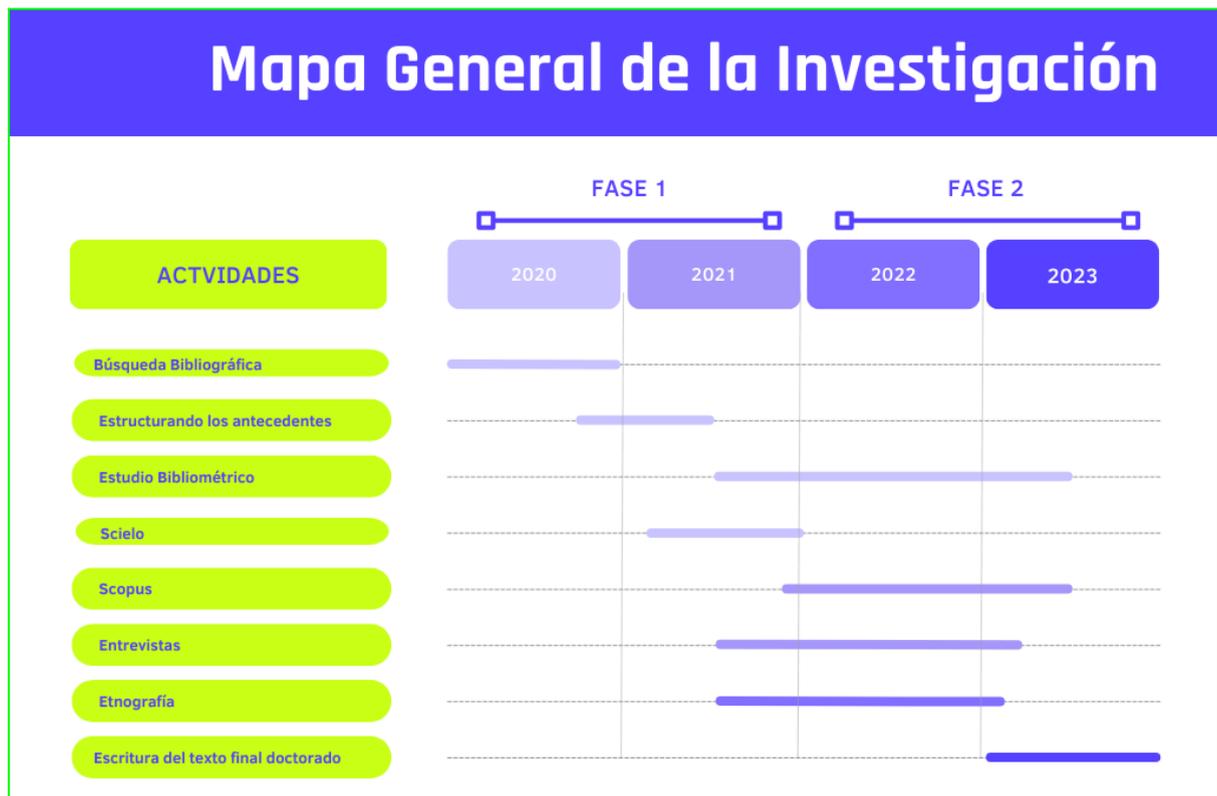
La Fase 2 incluyó un estudio etnográfico en el Laboratorio de Innovación (GobLab)<sup>3</sup> de la Universidad Adolfo Ibáñez, en Chile, entre agosto de 2021 y diciembre de 2022. En esta etapa, el análisis se centró en comprender cómo las prácticas científicas actuales reflejan características de la sociedad, con un enfoque en “evaluaciones sociales”.

---

<sup>3</sup> <https://goblab.uai.cl/>

Durante ambas fases, realicé más de 30 entrevistas con autores de artículos relevantes, ampliando la comprensión de contextos y desafíos en la investigación. A continuación, la Figura 1 presenta un esquema general de la investigación.

**Figura 1.** *Mapa general de la investigación*



La metodología sensitiva fue clave para permitir mayor flexibilidad en el proceso, como enfatiza Knorr-Cetina (1981), quien defiende un trabajo abierto a las indeterminaciones del campo. Este enfoque de “bricoleur”, inspirado en Knorr-Cetina, explora oportunidades y ajusta el recorrido según las circunstancias.

Es preciso decir que la elección de la incorporación de la metodología sensitiva como paraguas principal de mi esquema de investigación permite transitar por los distintos momentos de la investigación con mayor libertad y fluidez. Como diría la autora es preferible estar abierto a las indeterminaciones propias y provenientes de los contextos de investigación, que ser muy rígido en cuanto a las selecciones del trabajo científico<sup>4</sup>.

<sup>4</sup>Para saber más sobre la importancia de la indeterminación en contextos investigativos, ver Knorr-Cetina (2005), parte 9 del capítulo IV.

Considero entonces, que esta investigación está constituida como un *bricoleur oportunista*, parafraseando a la misma autora, no solamente por el ajuste constante en cuanto a los medios de recolección de información útil, como también por la capacidad de aprovechar las casualidades, encuentros fortuitos y redes dinámicamente construidas a mi alrededor, entendiéndolo como una ventaja investigativa.

Los *bricoleurs* son oportunistas. Son conscientes de las oportunidades materiales que encuentran en determinado lugar, y las explotan para lograr sus proyectos. Al mismo tiempo, reconocen lo que es factible, y ajustan o desarrollan sus proyectos según eso. Al hacerlo, están constantemente dedicados a producir y reproducir algún tipo de objeto factible que logre cumplir con el propósito que temporariamente se le ha asignado. [...] [Esos objetos] representan, no un producto perfecto de ingeniería, sino un *patchwork* de retazos puestos juntos cuando surge la oportunidad [...]. (Knor Cetina, 2005, p 113.)<sup>5</sup>

Podemos entender mi camino investigativo como un *patchwork* científico, una cobija hecha de retazos, de momentos investigativos no ordenados y algunas veces casuales. Donde a pesar de su tono flexible, se constituye alrededor de un objetivo eficiente y único, cuya función es cumplida.

Por otro lado, el estudio bibliométrico buscó mapear la producción académica en ciencias sociales y big data en América Latina, revelando temas, redes de colaboración y tendencias. Las categorías temáticas emergentes orientaron el análisis etnográfico, abarcando tópicos como ciencia social, big data y ciencia de datos en la región. La Figura 2 detalla estas categorías.

**Figura 2.** *Definición de las categorías de análisis.*

A= Observación de la interacción entre las ciencias sociales y los proyectos en ciencias de datos ejecutados en el laboratorio.

<sup>5</sup>Apuntes del cuaderno de campo, 19 de agosto 2021: “Leer esta parte me dio bastante alivio, en algún momento llegué a dudar de la científicidad de mis métodos. Hace algunos meses que vengo tratando de encontrar un modelo metodológico uniforme que se adecue a las circunstancias muy cambiantes de un campo con tanta innovación y novedad. También estoy pasando por desafíos y “coincidencias” laborales que me exigen cambiar constantemente la forma como había en un principio planteado la fase de ejecución de la investigación.

23 de agosto de 2021: “creo que estoy enamorada de Cetina! Sigo la lectura de donde había parado, me deparo con más esta noción de indeterminación del campo y estoy casi que me paro y aplaudo a esta mulher, mientras doy saltitos en la silla que estoy sentada! Esto es lo que está pasando conmigo! que felicidad saber que esto también es científico. Y que tengo permiso para seguir adaptando mis métodos, conforme me va saliendo el campo! Grifar bien esta página y usar como respuesta y base de sustentación metodología en la defensa de la tesis, caso me pregunten sobre las elecciones metodológicas del estudio.

B= Características cotidianas del trabajo.

C= Conceptos coordinados, relacionados y/o codependientes.

D= Determinar situaciones del campo de investigación latinoamericana.

E= Procesamiento computacional de los datos sociales

La fase de investigación bibliométrica fue alimentada por la búsqueda de los textos hospedados en las bases de datos Scopus y Scielo. Cada base de datos fue capturada y estudiada por diferentes razones y a través de diferentes mecanismos, según la disponibilidad de obtener de forma legal, información relevante en los textos y sus bases correspondientes<sup>6</sup>. Sin embargo, algunos de los criterios de búsqueda fueron estandarizados a título de mayor control de los resultados, aumentando así la uniformidad de recolección de información y posterior análisis. Aun sí en ambas bases de datos, los términos buscados eran “big data”; “social science”; “ciencias sociales/sociais” fueron examinados a partir de los títulos, las palabras claves y los resúmenes de las publicaciones.

Por fin, la elección de una perspectiva latinoamericana se justifica por la valoración de las especificidades locales en la producción de conocimiento, contribuyendo a una ciencia diversificada y consciente del contexto en que se produce. La elección de América Latina como foco de estudio se justifica, en esta investigación, en función de tres razones principales. Primero, el autor enfatiza la importancia de construir un conocimiento que refleje y valore las particularidades y fortalezas de la producción de saberes locales, especialmente en contraste con el dominio de las perspectivas del norte global en el ámbito académico.

Además, se inspira en la obra de Knorr-Cetina (2005), quien señala cómo las idiosincrasias locales contribuyen a diversificar la producción científica. Según esta perspectiva, el contexto

<sup>6</sup> El estudio bibliométrico en la base de datos Scielo fue conducido de manera autónoma. El estudio bibliométrico en Scopus fue realizado teniendo como base la metodología de investigación del CTS-Lab FLACSO Ecuador. Ver página web del laboratorio en: <http://ctslab.org/> <accedido en 27 de julio de 2022.

Mientras desarrollaba mis estudios, este laboratorio estaba presentando los resultados de su investigación en los congresos latinoamericanos de CTS. Por cuestiones académicas e intereses similares de investigación, me acerqué a unos de sus investigadores y pude acceder (en primera mano) a su “receta de análisis bibliométrico” antes que el artículo con los resultados finales de su estudio fuera oficialmente publicado. En términos bibliométricos la investigación del *CTS-lab* tenía los mismos objetivos que mi estudio, pero a escala global. Ellos tenían datos sobre las publicaciones en big data de todo el mundo, sin importar los diferentes marcadores entre las disciplinas. A mi me interesaba especialmente Latinoamérica. Así que decidí aplicar su modelo bibliométrico y construir conjuntamente con este investigador una versión similar latinoamericana a partir del análisis de la misma base de datos. Los resultados de este nuevo estudio conforma parte de los resultados presentados en el capítulo 7. Dejo la referencia del artículo base para que lo aprecien en su versión completa: Chavez, H., Albornoz, M. B., & Martín, F. (2022). ‘Big data’ Research: A Bibliometric Analysis of the Scopus Database, 2009–2019. *Journal of Scientometric Research*, 11(1), 64–78. <https://doi.org/10.5530/jscires.11.1.7>

local y las capacidades de cada laboratorio juegan un papel crucial en la creación de conocimiento, destacando cómo la ciencia está situada y condicionada por el tiempo y el espacio específicos:

“Una mirada de cerca a la escena de la investigación muestra que las selecciones son locales, y que dependen tanto del contexto de la investigación como de la situación concreta de investigación. Vemos las idiosincrasias implicadas en esas selecciones y cómo los criterios de decisión dependen más del proceso que de estipular (o gobernar) su cierre y su determinación. En resumen, ver de cerca la escena de la investigación nos obliga a volver a traer el tiempo y el espacio a las operaciones científicas y a concebirlas como operaciones localmente situadas” (Knorr-Cetina, 2005, p. 121).

Finalmente, porque se reconoce (Mignolo, 2008; Quijano, 2014; Domingues, 2015; Follari, 2015) que las capacidades de producción científica en América Latina están influenciadas por factores operativos, humanos y contextuales propios de la región. Esta “latinoamericanidad” es fundamental para entender cómo se genera y articula el conocimiento en estos países. En conjunto, estos factores reflejan el interés del autor en visibilizar una producción de saberes que es profundamente contextual y representa su lugar situacional como investigador en una tradición científica latinoamericana.

#### **1.4 Nota a los lectores: el trabajo de campo**

Antes de presentar los enfoques teóricos que inspiran esta investigación, me gustaría reafirmar el compromiso por crear conocimiento que hable de la construcción de la ciencia a partir de una mirada constructivista. Mi deseo por conocer la consolidación de los big data y el procesamiento computacional de los datos en el ambiente de investigación social, está anclado al supuesto de que los estudios de la ciencia, por más específicos que sean, deben considerar y reflexionar a partir de su propia imprenta científica. Como investigadora, no sólo se debe tener la función de recolectar y analizar datos objetivos sobre un mundo supuestamente natural, “hecho como es”, pues somos parte de la fabricación de este conocimiento que se produce en vivo, como diría Karin Knorr-Cetina, autora de mi libro académico de cabecera.

Considero, pues, que el presente texto es ante todo etnográfico. Me he basado en algunas etnografías de laboratorio, tales como, Latour y Woolgar (1995), autores con quienes uno contacta de inmediato; Michel Callon (1986) y Michael Mulkay (1979). Pero como ya he

dicho, la obra de Knorr - Cetina (1981) y su método sensitivo de construir los estudios antropológicos sobre la ciencia han sido mi guía.<sup>7</sup>

Reconozco mi lugar como científica, que pertenece y se influencia por conceptos pre-establecidos (o “selecciones”<sup>8</sup>, como dice la autora). Es decir, mi *background* teórico, mis intereses de investigación, la condición de mi trabajo, los recursos disponibles para tal y las casualidades propias de la vida, son partícipes en la construcción de mi propio saber científico.

Más que un norte epistemológico a seguir, considero que la postura constructivista de investigación es esencial, sobre todo para el campo de estudio que me propuse analizar. La intersección entre las ciencias de datos, ciencias computacionales y tecnologías de la información, a menudo son reconocidas como objetivas y neutrales, por muchos científicos. Optar por una postura constructivista en investigaciones de esta naturaleza significa, para mi, destacar la importancia de la reflexividad y la transparencia en la investigación científica en un entorno que muchas veces se pretende y se define como libre de pre-concepciones, construcciones y mandatos sociales.

Es parte del oficio del científico constructivista tomar consciencia de los propios intereses y selecciones, reconociendo la influencia que pueda tener en los resultados que se encuentren, independiente del campo de la ciencia en el cual trabaja.

Teniendo en cuenta esta perspectiva, considero este texto un bonito homenaje a los estudios de la ciencia y a la producción del conocimiento científico según las propuestas de Karin Knorr-Cetina (1981).<sup>9</sup> Su trabajo ha influenciado el mío no sólo por el contenido, también por su forma.

Entre tantos otros estudiosos de la ciencia, encontré una agradable y compatible manera de investigar, según lo que a mi me interesaba hacer y averiguar. Atribuyo esta decisión a nuestra mujeridad. En medio de tantos gurús de los estudios de la ciencia en las ciencias sociales, elegí a esta autora, mujer, como mi maestra y guía. Como discípula, traté de seguir

<sup>7</sup>Acá hago alusión al método sensitivo conceptualizado por Knorr-Cetina en su libro, versión traducida al español de 2005. Ver en Capítulo I, apartado 9.

<sup>8</sup>“Si los objetos científicos son selectivamente extraídos de la realidad, se los puede deconstruir, cuestionando las selecciones que incorporan; si estos son derivados de decisiones, pueden ser fabricados imponiendo decisiones alternativas. En la investigación científica, la selectividad de las selecciones incorporadas en el trabajo científico previo es en sí misma un *tema* para una nueva investigación científica. (Knorr-Cetina, 2005, p.63)

<sup>9</sup>Versión original: Knorr-Cetina, Karin. *Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science* 1° edición: Pergamon Press, 1981.

sus pasos, lo que me hace considerar esta investigación como un estudio de caso construido bajo sus preceptos (la forma de concebir y entender la ciencia) y sus enseñanzas (la manera de ejecutar un trabajo científico que estudia la ciencia).

Tal vez esta fijación por una fuente de inspiración principal suene casi limitada. Si bien puedo entender esta visión, y a riesgo de ser cuestionada por una falta de variedad de referencias bibliográficas, sostengo esta opción como una elección hecha a conciencia y principio.

Primero, porque en términos de material de referencia, Knorr-Cetina aparece en mi radar como indicación de lectura proveniente de uno de los evaluadores de la banca de cualificación en los comentarios hechos en el proyecto. Encontrar su texto digital y completo no fue difícil, como fue el caso de muchos otros textos clásicos de los estudios de ciencia que estaban en mi lista.<sup>10</sup>

Segundo, es una elección estratégica. El tipo de estudio que yo pretendía conducir y el contexto donde me encontraba encajaban perfectamente con su propuesta teórica, la epistemología por detrás de sus conceptos y con la forma que nos enseña a investigar la ciencia. Por estas razones, la obra de Knorr-Cetina fue la base y sostén teórico con el cual me apoyé desde la construcción del diseño de investigación a la manera cómo decido escribir este texto, relato del proceso de mi trabajo investigativo y esfuerzo científico propio.

Con respecto específicamente a la escritura de este informe de investigación, hay también un par de cosas que me gustaría aclarar: la propuesta constructiva de producción del conocimiento científico, combinado a la metodología sensitiva, permite y otorga licencia a una escritura en primera persona. Recurso que, ya han podido notar, uso sin excusas.

La propuesta contenida en este proyecto busca evidenciar los relatos de mi experiencia. En cuanto investigadora científica que se afecta, se aproxima al campo, hace inferencias y selecciones de acuerdo a sus capacidades y percepciones oportunas. No hay que esconder que

---

<sup>10</sup> Cuando conducía la búsqueda bibliográfica de mi investigación, estábamos todavía viviendo el periodo de transición de la digitalidad de los textos. El primer año del doctorado (2019) aun solíamos sentarnos en una biblioteca a leer un libro físico. Los textos académicos clásicos digitales aún no eran tan comunes en su versión completa. Había que andar repartiendo los libros para también no vulnerar los derechos de propiedad y venta de las editoriales y autores. Se leía digitalmente los artículos de revistas y publicaciones más recientes de los últimos 30 años. La introducción masiva de los textos (y clases y actividades evaluativas vía programas de reunión) se instauró a partir de la segunda mitad de 2020, en contexto de la pandemia global Covid-19. Estaba en mi segundo año de doctorado, tratando de leer todo lo que encontraba, pero la mayoría de las cosas estaban "picadas". Con las bibliotecas cerradas, solíamos hacer contrabando de textos y buscar Pdfs en grupos y foros de académicos sociales en internet. Nota de mi cuaderno de campo, abril 2021: "Encontrar el texto de Knorr-Cetina digitalizado y completo, fue un feliz caso de esas casualidades oportunas, o contingencias propias del campo, citando la propia Knorr-Cetina. La sensación que tengo es que no necesito otro raciocinio en que me espelhar".

los hallazgos que se presentan en este trabajo están contruidos en base a los datos e innumerables evidencias encontradas en conjunto a la estructura de mi propia jornada académica y profesional.

Gracias al sentirme cómoda en este lugar es que nace mi libertad de escribir de manera autorreferente y con un toque de lirismo. También tomé la libertad de mantener la escritura original de mi cuaderno de campo en algunas notas de pie de página. Se puede observar una mezcla de palabras en portugués y español, retrato de mi condición bilingüe. Aunque sin duda el español sea prioritariamente el idioma con el cual me acostumbré a escribir académicamente, los “accidentales” términos y palabras escritos en mi idioma madre, son una especie de resistencia literaria, necesaria y coherente con la propuesta investigativa en el marco decolonial. Una memoria o refugio de la manera más relajada con la cual escribo el portugués y que, por eso, me pareció interesante preservar.

Respecto a las notas de pie de página, me gustaría justificar la insistencia en su uso, pues son un intento estratégico de no esconder la presencia de las “indeterminaciones del hacer científico”, las “conexiones trans científicas”, o la “condición situada” de toda y cualquier investigación (Knorr-Cetina, 2005). Como nos elucida el esfuerzo del científico en no evidenciar las circunstancias incontrolables de su quehacer, acabó por constituir un práctica generalizada de “alienación de la investigación”, palabras mías, por parte de los científicos.

Una investigación alienada es aquella que esconde los procesos de producción del conocimiento. Que no trae a la luz, ni reconoce el papel fundamental de las coincidencias, de las sugerencias, de las comparaciones y de la influencia que la disponibilidad de recursos y capacidades determinan sus resultados. Por ello, en este texto, las notas de pie de página son una herramienta importante o una evidencia de los procesos de fabricación del trabajo científico. Contar los detalles, los bastidores de la investigación, los famosos “cahuines” de lo que pasa por detrás de las escrituras científicas es parte del propósito de, como investigadora, no alienarme en la confección de mi propio saber. Es una forma de llamar la atención hacia el trabajo humano que está presente, y muy vivo, detrás de la producción de conocimiento de los textos académicos.

Espero que como un “buen cauhín” (en chileno), o en portugués “fofoca”, estas notas sean también una pausa refrescante a esta lectura tan técnica. Definitivamente su contenido contiene evidencias muy valiosas de que incluso en la era de lo automáticamente artificial,

existe decisión y determinación humana por detrás de las letras, números, tablas y figuras discadas en los dispositivos digitales. Esta es una de las bellezas de las investigaciones constructivistas y la fortaleza del texto etnográfico: permite ciertas libertades, sin dejar de ser constitutivamente racional y científico. Espero que más que informativo y esclarecedor, este texto pueda ser un disfrute.

Otro aspecto relacionado a la escritura de este texto es la disposición y representación anacrónica de las evidencias, ejemplos y casos expuestos en este documento. Mi mayor preocupación está en demostrar los aspectos encontrados a través de una uniformidad temática y de raciocinio, más que disponerlas en orden temporal. La idea de la disposición de los capítulos es facilitar la comprensión del lector, más que explicitar las evidencias de forma cronológica, o la secuencia exacta de los acontecimientos. Sin embargo, un detalle: aunque el periodo de investigación etnográfica “oficial” este calendarizado hasta final de diciembre de 2022, a la actualidad, sigo involucrada con algunas actividades del laboratorio estudiado y participando en distintos espacios de formación y redes de investigadores en ciencia de datos e inteligencia artificial, que también fueron fuente de material de análisis o evidencias, incluso en este mismo momento, en que estoy dedicada exclusivamente a la escritura de este proyecto.

En otras palabras, he recolectado material hasta el momento de escribir este texto, en 2023. En mi opinión, esta es una de las fortalezas del paradigma constructivista de los estudios de la ciencia: la dádiva de participar de estudios dinámicos, que se transforman y actualizan desde del laboratorio hasta el momento de la escritura de los informes de investigación.<sup>11</sup> Como más o menos diría Karin Knorr-Cetina, 2005, el proceso de construcción del conocimiento es el premio resultado de la experiencia viva que la atribución de sentido a los datos observados proporciona a lo científico.<sup>12</sup> Espero haber logrado un buen uso de los regalos atemporales que he recibido a lo largo de este trabajo de investigación.

Hablando de regalos, me considero afortunada con respecto a las experiencias y puertas que se fueron abriendo mientras estudiaba el tema de los macrodatos en las ciencias sociales. Considero haber sabido aprovechar las oportunidades de las *conexiones transcientíficas*<sup>13</sup> que

---

<sup>11</sup>Según Knorr-Cetina (2005), los textos académicos y los artículos científicos son también construidos y presentados de manera interesada, contingencial y estructurada con el fin de presentar un conocimiento o resultado de manera objetiva pero también interesada. Ver capítulo V del libro

<sup>12</sup>“No debería sorprender, entonces, que los científicos estén habituados a los premios inesperados que la experiencia viva les puede dar en este proceso de construcción de sentido, un beneficio que algunos científicos sociales parecen haber olvidado” (Knorr-Cetina, 2005, p. 313).

<sup>13</sup> (Knorr-Cetina, 2005, p. 218 en adelante).

se fueron presentando, y que hoy se traduce en dos grandes implicaciones con respecto a la artesanía de mi propia investigación.

La primera, la misma definición metodológica de trabajo a seguir: con el objetivo de ir adaptando la recolección de información durante la investigación, utilicé variados recursos metodológicos. Esta estrategia hace parte del sustento antropológico que abriga la metodología sensitiva, lo que me permitió elegir con fluidez y libertad las técnicas de recolección de información más adecuada según cada momento de la investigación.

Conduje un estudio bibliométrico, realicé entrevistas, análisis de documento, observación participante, actué como consultora de investigación, estuve en terrenos conduciendo una etnografía y mantuve mi cuaderno de campo. Paralelamente, combinaba, variaba o intercambiaba las estrategias metodológicas a utilizar, sin perder de vista mi destino final ni la pregunta que me guiaba. De esta manera, fui adaptando técnicas de recolección de información según las contingencias y especificidades de cada momento vivido en las diferentes fases de la investigación.

Lo que nos lleva al segundo punto: la condición contextual de la producción de los saberes acá presentados. La producción del conocimiento que presento forma parte de un entendimiento de que mi propia y específica trayectoria académica es determinante en la generación de los resultados a exponer. No es posible desvincular la contingencia, ni el propio interés de investigación. La misma elección de las bases teóricas y conceptuales que sustentan este estudio determinan y provocan, en gran medida, las decisiones que he tomado en la tentativa de analizar y decodificar las informaciones y datos con los cuales he trabajado.<sup>14</sup> Sostengo que la construcción del saber presentado tiene justamente valor por esto.

Teniendo consciencia de que los resultados de esta investigación son resultado de un proceso de fabricación dependiente y estructurado de acuerdo a mi realidad, mi interés y según los recursos disponibles en este momento, no puedo dejar de mencionar que mi posición e interés

---

<sup>14</sup>Sobre la importancia de contextos en la fabricación del conocimiento científico, ver: “La tesis que estamos considerando es la de que los productos de la ciencia son construcciones contextualmente específicas que llevan las marcas de la contingencia situacional y de la estructura de intereses del proceso por el cual son generados, y que no pueden ser comprendidos adecuadamente sin un análisis de su construcción. Esto significa que lo que ocurre en el proceso de construcción *no* es irrelevante para los productos que obtenemos. También significa ver los productos de la ciencia como *sumamente estructurados internamente* mediante los procesos de producción, independientemente de la cuestión de su estructuración externa por la vía de alguna coincidencia o no coincidencia con la realidad”. (Knorr-Cetina, 2005, p. 61)

siendo una científica, que observa de cerca un laboratorio específico de un país latinoamericano del sur global, me llevan, también, a resultados situados.

No resta duda declarar, entonces, que vivir el tiempo digital e informacional que nos encontramos, es parte constituyente de los hallazgos de la investigación. Es por esto que el énfasis en la digitalidad de la vida y los datos es un tema que orbita la discusión presentada en esta investigación, aunque no sea el tema central.

Investigar en contexto de digitalidad implica un desafío que atraviesa todo el contenido de este trabajo y significa un desafío pues todavía estoy aprendiendo a manipular y hacer uso de “adaptaciones chuecas y no óptimas” (parafraseando a un entrevistado) de los innumerables sistemas de recolección, categorización y análisis de datos digitales que no paran de surgir con la popularización de la inteligencia artificial.

Como mencioné, mi contacto con temas tecnológicos no pasaba de la discusión sobre el desarrollo de las técnicas y tecnología científicas y su impacto en la sociedad. Adentrarse verdaderamente en el tema de los datos masivos y la ciencias de datos, requirió un entratamiento no menor de mi parte. Otro desafío respecto a la necesidad de adaptación de antiguas o viejas conocidas técnicas de investigación social a los contextos digitales ha sido: ¿Cómo conducir una etnografía en un escenario de reuniones online? ¿Cómo absorber información relevante a través de la observación como participante digital? ¿Cómo sentir el campo si lo único que veo son rostros sin mucha expresión corporal, sin instancias de cafés y conversaciones de pasillo? ¿Sobre qué etiqueta declaro que analicé los infinitos pantallazos de emails, páginas web, reuniones, seminarios y conferencias en internet? La participación, 1 año después, en un evento grabado y disponible en el canal de youtube de un congreso de mi área de estudio, ¿puede ser considerado como participación y consecuente material de campo?

La emergencia de estas novedades en relación a la manera de obtener información válida y relevante, científicamente hablando, posibilita también la creatividad con respecto a los análisis, presentación o difusión de estos mismos conocimientos. No solamente las técnicas y maneras de manipular los datos han pasado por un proceso de remodelación. La forma como se presentan los datos y la necesidad de que estos sean cada vez mejor entendidos, hace que científicos del momento presente, dispongan de propuestas de visualización de datos tampoco

convencionales. “Es necesario elevar la calidad de la información entregada. Y esto, la tecnología nos lo ha permitido mejorar”, apuntó una de las académicas entrevistadas.

Así que el reporte de este estudio está lleno de recursos digitales que incrementan la presentación de los resultados. Por esto, se recomienda leerla de manera digital, si es posible. Más que ser un texto académico, relativamente largo y aburrido, este trabajo convida a una experiencia para las afueras del texto mismo. Hay diversas tablas, gráficos y mapas que son interactivos, posibilitando al lector una mayor autonomía para observar la producción de información que proviene de esta investigación.

Finalmente, espero que las reflexiones que siguen sirvan de inspiración para el científico social interesado en la digitalidad de los datos, el procesamiento computacional, y la utilización de macrodatos en las investigaciones sociales. Más que ilustrar un panorama emergente con respecto a la fabricación de nuestro propio conocimiento social y científico, espero levantar un debate con respecto a la oportunidad de repensar la práctica investigativa en las ciencias sociales, a fin de reconocer a través de estos signos, el desarrollo contemporáneo de la sociedad y su forma de existir.

En la parte 2 de este informe, se presentan los principales resultados encontrados en la investigación. Demostraré los hallazgos relacionados al estudio bibliométrico para dar a conocer el panorama general de la investigaciones sociales en big data en latinoamérica y a modo de comparación, también en el mundo. En la parte 3, trabajaré con los hallazgos provenientes del momento etnográfico de la investigación, conducida por 1 año y 4 meses, de la participación en un laboratorio de innovación en una prestigiosa universidad chilena. Discutiré las implicaciones sociales de los estudios en ciencias sociales que utilizan conocimiento, tecnología y macrodatos.

Después, en el capítulo 10, presentaré las conclusiones con respecto a los hallazgos de la investigación, en un esfuerzo y cierto atrevimiento en sugerir, según el análisis realizado, el futuro del trabajo del investigador social contemporáneo interesado en incorporar la ciencias de datos y los macrodatos a los trabajos de investigación social.

## Capítulo 2: Discusión teórica: lentes para el campo

### 2.1 El fenómeno *big data*

El término *big data*, original en inglés, y hoy adoptado por el español, es, también, conocido como macrodatos, datos masivos, inteligencia de datos, o datos a gran escala. “Es un término que hace referencia a conjuntos de datos tan grandes (masivos) y complejos (diversos) que precisan de aplicaciones informáticas no tradicionales de procesamiento de datos para tratarlos adecuadamente” (Lupton, 2015). El término ha estado en uso desde la segunda mitad de los años 90, haciendo principalmente referencia a la cantidad de datos que superaría la capacidad de un software convencional para ser capturado, procesado y analizado en un tiempo razonable (Mashey, 1998).

Aunque su característica más evidente (ya desde el uso de “big” en el concepto), sea la cantidad de datos que contiene, no es el orden de grandeza lo que determina la inclusión de una base de datos en la categoría de big data. Hay una controversia importante sobre las dimensiones mínimas de cuán grande podría ser el dato big data.

El orden de grandeza de los macrodatos varía dependiendo del contexto y de la aplicación específica. Generalmente, se considera que la condición para determinar si un conjunto de datos es o no *big*, es que sea de un tamaño de por lo menos terabytes ( $10^{12}$  bytes) y pueden llegar hasta petabytes ( $10^{15}$  bytes), exabytes ( $10^{18}$  bytes), zettabytes ( $10^{21}$  bytes) e incluso yottabytes ( $10^{24}$  bytes)<sup>15</sup>.

Sin embargo, no importa tanto el tamaño de los datos sino su complejidad y heterogeneidad, así como la necesidad de utilizar técnicas y herramientas específicas para su procesamiento y análisis. Se compone de diversas fuentes de información, incluyendo datos generados por usuarios en redes sociales, transacciones financieras, sensores, registros médicos y otras fuentes de información en línea y fuera de línea. Esta información se presenta en diferentes formatos, como texto, imagen, audio y vídeo, al paso que van dejando registro de las más variadas actividades.

---

<sup>15</sup>Arthur, Charles (29 de junio de 2011). “What's a Zettabyte?” By 2015, the internet will know, says Cisco». *The Guardian*. Accedido en marzo de 2020

**Figura 3. Tipos y tamaños big data**



*Nota: Henry Chávez, imagen presentación Seminario Académico, Núcleo Especulativo sobre la Tecnociencia y los Biomateriales, 2022, Chile.*

Las huellas digitales, o rastros que cada usuario va dejando en su vida digital, se van acumulando en el ambiente de la nube y algunos pocos sistemas propios de almacenamiento (Rocha, 2018). Ambos están respaldados en supercomputadores distribuidos de forma dispersa en cada rincón del mundo, bajo el control y en manos de dos o tres empresas de tecnología de la información y comunicación.

A partir de los 2010, la definición de *big data* se relaciona directamente con los activos de información de este conjunto de datos. Es el conocido 3V de los grandes datos: volumen, velocidad y variedad. Pero con el pasar de los años y su mayor utilización y comprensión, las características de los datos masivos se expresan, más contemporáneamente, a partir de las 5 V (Kitchin, 2014):

1. Volumen: la cantidad de datos generados y guardados.
2. Variedad: el tipo y naturaleza de los datos: textos, imágenes, audio y video.
3. Velocidad: enmarcada en el contexto y desafíos de su análisis, a partir de cumplimiento de las exigencias de cada proyecto.
4. Veracidad: se refiere a la calidad de los datos capturados, que pueden variar mucho y por tanto afectar los resultados del análisis.
5. Valor: los datos generados deben ser útiles, accionables y constituirse en un activo.

Para procesar macrodatos se requieren herramientas y técnicas específicas. Por ejemplo, algoritmos de aprendizaje automático, análisis de redes sociales, minería de datos<sup>16</sup>, procesamiento de lenguaje natural, entre otras técnicas de análisis de datos, combinadas, posibilitan la identificación de patrones y tendencias a una escala nunca antes experimentada, ya que hasta antes había una limitación tecnológica importante en cuanto a capacidad de procesamiento y almacenamiento de los computadores. En la práctica, procesar datos masivos, es utilizar técnicas avanzadas de procesamiento, como la minería de datos, el aprendizaje automático y la inteligencia artificial, para extraer información útil de un mar infinito de informaciones.

Precisamente por ello, y por prometer facilidad y eficiencia a la hora de procesar y analizar los resultados de lo que genera, es que el uso de macrodatos fue ganando cada vez más espacio e interés en el mundo científico en investigaciones, proyectos de investigación y laboratorios de diversos campos, incluyendo las ciencias sociales. Pero más allá de la capacidad de optimizar algunos procesos de investigación computacionalmente, la utilización de/a través de big data implica una serie de características y desdoblamientos con respecto al quehacer científico involucrado.

Como se puede notar, la incorporación de la veracidad y el valor a las características necesarias de big data revela un entendimiento fundamental sobre los desafíos de aplicabilidad y tratamiento de un conjunto de datos de gran magnitud, y que hacen todo el sentido e importancia a la hora de pensar los usos de big data en la ciencia: no basta con tener a disposición el máximo posible de información sobre lo que a uno le interesa. Uno de los desafíos más particulares de la gestión del big data es justamente el buen manejo de los mecanismos, estrategias y aprendizajes para ser implementado en todas las etapas de su manipulación (captura, almacenamiento, minería y procesamiento), hasta llegar por fin a su análisis (Rocha, 2018).

---

<sup>16</sup>La minería de datos, también conocida como descubrimiento de conocimiento en bases de datos (KDD, por sus siglas en inglés), es un proceso de exploración y análisis de grandes conjuntos de datos con el objetivo de descubrir patrones, tendencias y relaciones útiles. La minería de datos utiliza técnicas y algoritmos de aprendizaje automático, estadística, visualización de datos y otras disciplinas relacionadas para analizar datos estructurados, semiestructurados y no estructurados. Estos datos pueden provenir de diversas fuentes, como bases de datos empresariales, registros de transacciones, redes sociales, sensores, registros web y poseen aplicaciones en una amplia gama de campos, como el comercio electrónico, la salud, las finanzas, el marketing, la seguridad, la investigación científica, etc. El proceso de minería de datos generalmente implica los siguientes pasos: definición del problema; recolección de datos; preprocesamiento de datos; selección de características; aplicación de algoritmos de minería de datos, interpretación y evaluación de resultados; implementación y seguimiento. Definición de Jiawei Han en *Data Mining: Concepts and Techniques* (2011).

Lo que se denomina “condición de big data”<sup>17</sup>, sin embargo, no debería estar pautada solamente por el tamaño y la velocidad de ellos, lo que no pasaría de una simple descripción de las bases de datos. De acuerdo con la postura crítica de las ciencias sociales, el análisis hecho con y a través de los macrodatos, como por ejemplo el aprendizaje automatizado o la minería de datos, permite encontrar patrones y modelos útiles para un entendimiento de lo social, de forma alternativa a los levantados hasta hace muy poco tiempo (Ribes, 2018).

La digitalidad de los datos y la introducción de técnicas computacionales en el proceso investigativo, sobre todo en las ciencias sociales, tradicionalmente alejada de estos mecanismos de investigación, empuja una discusión respecto de su utilización y consecuencia, en la producción del conocimiento científico. Cabe además preguntarse, si reestructura o si se retroalimenta con nuevas referencias su propio quehacer, y cuáles son las posibilidades de fabricar sus propios saberes.

Es por esto que defiendo la importancia de reflexionar sobre las implicaciones científicas y discursivas con respecto a la implementación de big data, del procesamiento de datos computacionales y de las herramientas de procesamiento automático de datos al interior de nuestra disciplina. Los grandes conjuntos de datos pueden proporcionar información detallada sobre las tendencias y patrones en la sociedad, lo que permite a los investigadores realizar análisis con niveles de comparación y complejidad diferentes de lo que se venía trabajando.

## 2.2 El (big) dato digital

Como sociedad hemos sido capaces, a través de la tecnología e internet, de proyectar el mundo material en el espacio virtual. Construimos un ambiente digital similar a nuestra presencia en el mundo natural que, sumado a las herramientas tecnológicas, permite no solamente la producción de una cantidad inimaginable de datos, sino también una capacidad de almacenamiento de estos cada vez mayor. Con ello, experimentamos un incremento significativo de técnicas, modelos, aparatos y estrategias de producción, captura,

---

<sup>17</sup>Este término surge a partir de mi experiencia etnográfica, especialmente durante las entrevistas, frente a la pregunta si los científicos consideraban su banco de datos como “big data”. En mi diario de campo puede leerse una nota que dice: “Quanto mais entrevistas eu faço, mas eu percebo a mudança do meu vocabulário. Agora me sinto mais segura para fazer perguntas técnicas. Já não pergunto mais “que tao big é o seu bando de dados”, as vezes sim, mas geralmente complemento: para você, o que faz que o banco de dados que você está trabalhando seja -BIG- data? Existe m seu ponto de vista”. (Diario de campo, septiembre, 2022).

procesamiento y análisis de datos que, al fin y al cabo, forman parte del modo en que se interpreta el mundo.

Es en la digitalidad de la vida social y de las prácticas humanas, y en el estrechamiento de su relación con los aparatos tecnológicos, que se da la intensificación de la producción, captura, procesamiento y análisis de los datos (Nascimento, 2016). Desde esta realidad emerge, entonces, el big data: al buscar cómo digerir la impresionante cantidad de datos que producimos a diario, la cual se genera en redes y gracias al masivo uso dispositivos tecnológicos-digitales en nuestra cotidianidad, y que son el insumo a través del cual se movilizan la economía y la sociabilidad (Rocha, 2018).

Big data está, entonces, indiscutiblemente vinculado a la digitalidad de los datos y de la información, participe del contexto de digitalización de la vida social contemporánea. De hecho, su origen nace de la capacidad computacional y de almacenamiento posibilitada por las plataformas digitales, después del *word wide web* (*www.*), en los 90. El crecimiento de la producción y almacenamientos de datos en los medios digitales impulsó innovaciones tecnológicas capaces de soportar y procesar tantos datos (Chavéz, 2022). Así que también está asociado a la capacidad de distribución de la gestión, procesamiento, y almacenamiento de los datos en varios ordenadores, lo que ocasionó una verdadera revolución en términos de capacidad de análisis de una gran cantidad de datos.

“Debido a las grandes transformaciones producidas por la aplicación de esta nueva tecnología, la noción de *big data* comenzó a extenderse desde la ingeniería informática y los campo de datos de otras áreas con inmenso potencial como la salud, los negocios, las finanzas y las ciencias sociales” (Chavéz et al, 2022, p.25).

La indisociable relación entre macrodatos, digitalidad y la capacidad computacional e informática involucrada en su uso viene impactando la manera en que fabricamos conocimientos en materia social. Más específicamente, en cómo podemos incidir en la sociedad a través de nuestras investigaciones y estudios.

A pesar de la brecha digital que deja fuera a 51.8% de los habitantes del mundo, el volumen, variedad y velocidad de datos que pueden ser sociológicamente analizados es bastante grande. Bastaría con recoger las cifras relacionadas a las redes sociales digitales como

Facebook que sólo en el año 2016 tenía 1090 millones de usuarios activos diarios. Sumado a esto, están las informaciones vinculadas a celulares, imágenes satelitales y las transacciones de los sectores público y privado que, juntas y procesadas de la manera apropiada, podrían ser fuentes de información inagotable para el entendimiento de los fenómenos sociales experimentados por una sociedad cada vez más digital e interconectada por medio del internet (Ballivian, 2016).

Big data, por lo tanto, toma el registro del programa de investigación de la inteligencia artificial de los años 50/60, sumado a la explosión y la capacidad que tiene el internet de procesar información, para modelar y codificar unos cuantos datos que, reunidos, pretenden constituirse en información relevante sobre las conductas sociales humanas a partir de su comportamiento en la digitalidad. Queda claro entonces, que la profundización de las problemáticas que involucran al big data y la ciencia atraviesan una vía de reestructuración pragmática del trabajo investigativo práctico, metodológico y del saber epistemológico construido en torno a la disciplina, sobre todo en las ciencias sociales.

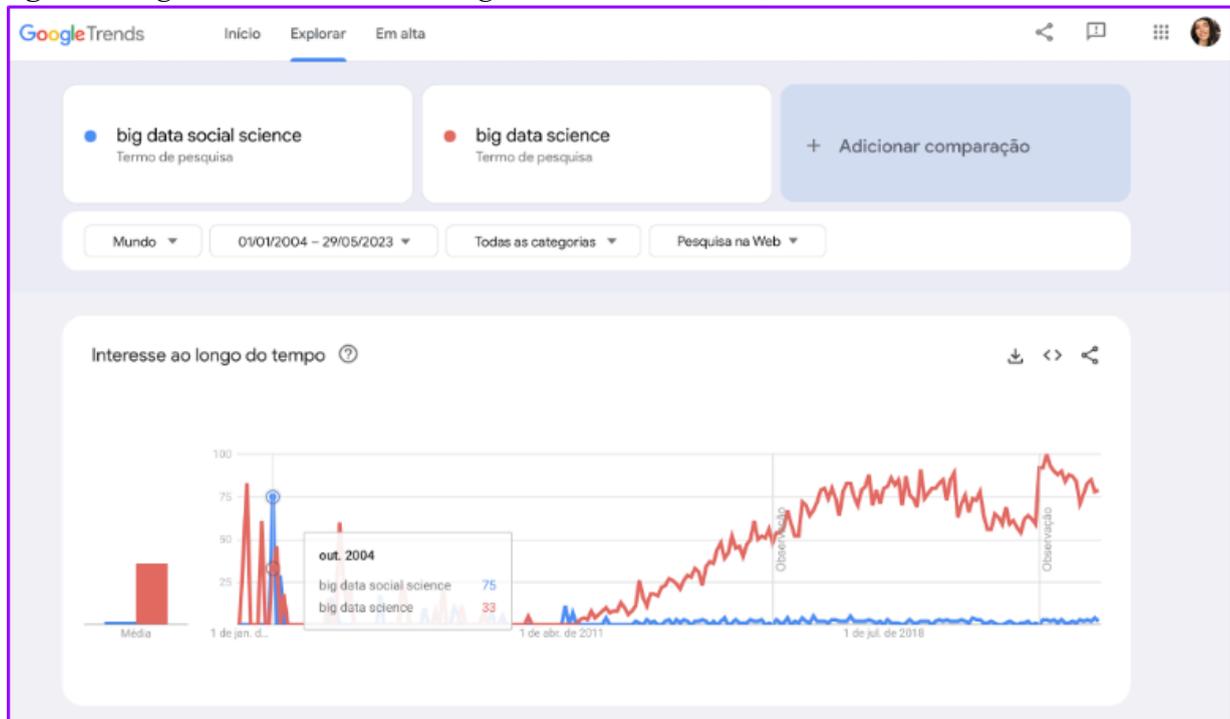
### **2.3 Big data y las investigaciones en ciencias sociales**

La utilización de macrodatos ha sido exitosa debido a la capacidad de obtención de informaciones sobre los individuos por medio de las huellas digitales que los usuarios de internet y dispositivos digitales van dejando en su paso por las redes. Por una cantidad relativamente pequeña de inversión, una empresa, por ejemplo, puede poseer datos muy específicos sobre individuos en particular, y al mismo tiempo, dado al flujo de las informaciones y el procesamiento computacional, agrupar y transformar esta información específica en nuevos conjuntos de datos agregados, de forma escalable. Esta es una de las maneras como los mercados de diversos países han predecido el comportamiento de ciertos grupos sociales, en base a sus patrones de compra y preferencias, consolidando y probando variedades de modelos previsibles. Hoy, a través de la inteligencia, lectura y procesamiento de la información y datos que dejan los consumidores, el mercado ha cambiado de manera única el modo en que funcionan sus negocios.

La ciencia no tardó mucho en incorporar a sus procesos de investigación tipos similares de tratamiento y análisis de datos. Si los macrodatos son utilizados para estudiar más exhaustivamente los hábitos de consumo de grupos enteros, bien podrían aplicarse en los más variados campos de la ciencia y las ciencias sociales (Navarro & Salazar, 2007).

En 2004, la tendencia en incorporar el análisis de los datos masivos en las investigaciones y problemáticas científicas se intensifica. La búsqueda del término “big data science” aparece por primera vez en internet ese año<sup>18</sup>.

**Figura 4.** *Big Data Science versus Big Data Social Science, de 2004 a 2023.*



Nota. Elaborado por Google Trends.

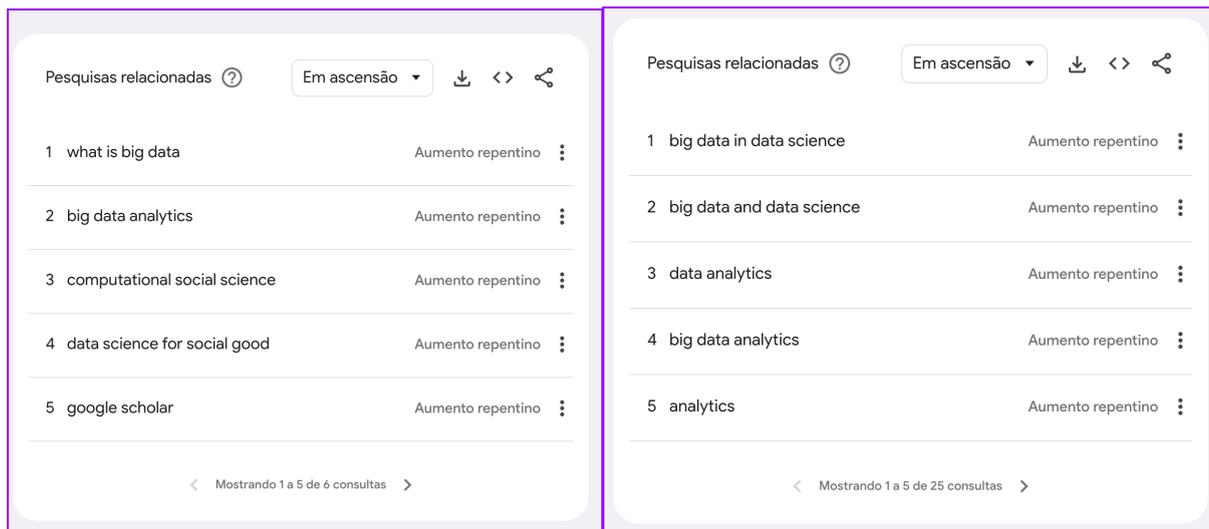
Tras una década de un tímido interés, en 2014 vuelve a observarse un boom de búsquedas del mismo término, periodo que coincide con la creación de las importantes revistas académicas que versan sobre *big data* en el mundo: *Journal of Big Data* (2014), *Big Data Research* (2016), *International Journal of Data Science and Analytics* (2016)<sup>19</sup>.

<sup>18</sup> Google Trends:

<https://trends.google.com/trends/explore?date=all&q=big%20data%20social%20science, big%20data%20science> e accedido en 20 de junio de 2022.

<sup>19</sup><https://journalofbigdata.springeropen.com>; <https://www.journals.elsevier.com/big-data-research>; <https://www.springer.com/journal/41060>. La antigüedad de estas revistas no eran siempre declaradas. Para obtener esta información revisé el año de la primera publicación de la revista en cuestión.

**Figura 5.** *Términos de la búsqueda repentina en 2014*



*Nota.* Según Google Trends.

La incorporación de los datos masivos digitales en el ámbito científico aporta a la ciencia la generación de nuevos repertorios conceptuales, nuevas estrategias y procedimientos investigativos, lo que incide a su vez, en la generación de nuevas habilidades científicas. Cambia la manera en que se comunica la ciencia y la distribución geográfica del trabajo y las plataformas de colaboración se amplían y diversifican, por citar algunos ejemplos (Rocha, 2018). Según la autora, dicho movimiento estimularía el aprendizaje de nuevos conocimientos, habilidades y lenguajes por parte de los científicos, que tienden a asociarse cada vez más con los profesionales de computación.

El aprendizaje automático y los algoritmos, han despertado cada vez más el interés de innumerables investigadores, bajo la promesa de practicidad, rapidez, precisión, objetividad y replicabilidad en sus estudios (Pérez, 2002).

Sin duda,

“El lenguaje de la programación modifica las estrategias de investigación y las formas de trabajo, lo cual no es tarea fácil para los campos disciplinarios que se configuran desde un corpus teórico y métodos específicos (Benoit y Cukier, 2015). Se podría decir que a partir de los grandes datos, presenciamos un momento en el cual se abren posibilidades para la imaginación sociológica, como la entendió Charles Wright Mills (2003), ya que amplía las posibilidades de observación, experimentación, y abre la puerta a una nueva practicidad”. (Rocha, 2018, p. 416).

Del deseo de comprender mejor los objetos y sujetos de estudio, emerge la configuración de nuevos campos disciplinarios como la ciencia de datos y las ciencias sociales computacionales (Conte et al., 2012). Desde este lugar, se reafirma el papel de las ciencias sociales de entender, dada la amplia inserción de los datos masivos en la vida social, los fenómenos sociales complejos y globales, tales como estudios de salud, población y epidemias en diversos continentes. Por otra parte, aporta a las investigaciones que utilizan datos de geolocalización para estudiar movimientos migratorios, desastres ambientales y otras causas humanitarias; abre nuevas preguntas e hipótesis desde la sociología política aplicada a las campañas electorales y preferencia de votos, auditoria popular, gestión de consejos; finalmente, gestión de las grandes ciudades, *smart cities* y el control de la urbanización, entre otros ejemplos.

Aunque el término “big data” sea cada vez más utilizado en la cotidianeidad y en espacios sociales fuera del mundo corporativo y científico, todavía está en construcción conceptual para las ciencias sociales: más que disfrutar de las características, potencialidades y promesas que la manipulación de los macrodatos posibilita al desarrollo de la ciencia como un todo, las ciencias sociales computacionales vienen consolidándose como un campo que reflexiona, cuestiona y propone una comprensión analítica de las implicaciones epistemológicas y metodológicas que el uso de los datos masivos puede generar a la producción del conocimiento científico y a los saberes sociales (Bastin & Tubaro, 2018; Rocha, 2018; Nascimento, 2016).

Así, se vuelve fundamental comprender el rol analítico que la utilización de los datos masivos desarrolla en la academia. Como apunta Boyed & Crawford (2011), “es crucial hacer preguntas sobre los supuestos analíticos, los marcos metodológicos y los sesgos subyacentes al fenómeno” (p. 2). Eso quiere decir que big data no sólo refiere a conjuntos de datos muy grandes y a las herramientas y procedimientos utilizados para manipular y analizar grandes cantidades de datos, sino también implica un giro computacional en la manera de pensar la

investigación (Burkholder, 1992), “*Big data* ha surgido como un sistema de conocimiento que ya está cambiando los objetos del conocimiento, al tiempo que tiene el poder de informar como entendemos las redes humanas y la comunidad” (Boyed & Crawford, 2011, p.3).

Por otro lado, se torna innegable que la manera de manipular los datos *bigdatamente* puede aportar al conocimiento de fenómenos sociales complejos. Dada su potencialidad de permitir la observación de un gran contingente de variables digitales, transforma la estructura de los datos que, al ser interpretados y analizados conjuntamente, aportarían a una comprensión de la sociedad diferente de lo que se venía haciendo desde entonces. Una de las promesas difundidas por los científicos sociales computacionales en la introducción de los macrodatos a las ciencias sociales es, precisamente, el mejoramiento de la toma de decisiones a través de la elaboración de modelos, correlaciones y proyecciones. Según el *Manifiesto of Computational Social Science*, publicado en 2012, de Conte, R. et al, uno de los mayores desafíos enfrentados por las ciencias sociales contemporáneas es la ejecución de estudios que aborden grandes problemas de ámbito social.

“How can we understand and help manage complex social systems if we still do not understand the basics of sociality? Is the new world of automated information treatment going to provide any help? The answer is: yes, ICT can provide significant help for social science. Not only ICT can help access, analyse and build upon BigData, i.e. new type of massive data, for addressing BigProblems. It can also help provide instruments for BigThinking. Indeed, computational social science can be characterised along two main aspects, which both take advantage, one way or the other, from ICT developments: a) BigData, and b) the role of computation in inspiring, formalizing and implementing the core scientific concepts, principles, and ideas of computational social science.” (Conte et al, 2012, p.331).

Es por esto que en este trabajo se plantea que la introducción de los macrodatos a las investigaciones en ciencias sociales implica un cambio en la manera de entender las sociedades, y una reflexión sobre la sociedad gracias a la especificidad que entrega investigar “a través” y “por la naturaleza” del big data.

La recolección masiva de datos digitales ha logrado producir una gran cantidad y variedad de flujos y conexiones entre datos. Estos, operacionalizados en las plataformas de manejo computacional de datos digitales y aliados a las tecnologías de inteligencia artificial,

prometen ser capaces de predecir, formalizar y conocer los comportamientos humanos en sus más variadas dimensiones: desde la preferencia de voto y de consumo, al comportamiento de masas y la movilidad urbana, como ya se ha observado. Sin embargo, la comprensión de la complejidad del pensamiento humano y su reflexividad no debería ser estudiada exclusivamente bajo la conjunción de estas técnicas de procesamiento de datos. Los matices y especificidades propias de la reflexividad y de la adaptación de los humanos siguen siendo una tarea analítica y artesanal, incluso en los ambientes donde los procedimientos computacionales ya están diseminados.

Independientemente de los resultados que representa la introducción de los macrodatos en la investigación social, aún deben probarse los límites y las potencialidades de su uso. Evaluar cómo las metodologías digitales de análisis de la sociedad están impactando el modo en que hacemos sociología es una preocupación no menor. En otras palabras, interesa saber qué efecto está teniendo la manera de utilizar y trabajar “bigdatamente” la información social, en la idea de las sociedades futuras, a través de la trasmisión de saberes constituidos al interior de las ciencias, y cómo los científicos sociales se apropian de estos para construir científicamente una representación sobre lo social y/o la sociedad.

La introducción de los macrodatos en las investigaciones en las ciencias sociales viene a cambiar los niveles de epistemología y ética de los trabajos que de ellos hacen uso: replantea las preguntas claves sobre la construcción del conocimiento, los procesos de investigación, y el cómo se comprometen con la información producida. El big data replantea, entonces, nuevos terrenos de objeto, métodos de conocimiento y definiciones de la vida social, ofreciendo a las disciplinas humanísticas una nueva forma de reclamar el estado de la ciencia cuantitativa y el método objetivo (Boyd & Crawford, 2011).

Se considera fundamental, por lo tanto, investigar la problemática big data de los estudios en ciencias sociales bajo una perspectiva que reflexione sobre el impacto que la producción de un conocimiento tiene en los resultados de su trabajo. Todo conocimiento científico está vinculado al contexto socio técnico en el que los datos son generados (Bunge, 1997). Esto explica el interés por entender más allá de lo que está siendo dicho por los científicos sociales respecto al big data.

Este estudio propone y presenta, de igual manera, resultados provenientes de la observación del quehacer científico construido en su cotidiano. La observación, categorización y análisis

de la artesanía y el caminar de los científicos y sus núcleos, laboratorios y proyectos de investigación, es el mayor aporte que proporcionará dicha investigación.

En resumen, esta investigación se justifica por la necesidad de comprender y analizar el uso de big data en la investigación social en Latinoamérica, y entender, en alguna medida, su impacto en la construcción del conocimiento social. Aspira, por otra parte, contribuir al desarrollo del enfoque metodológico y teórico de las ciencias sociales contemporáneas y latinoamericanas en la era digital.

## **2.4 La investigación científica social**

La emergencia del uso de big data en el contexto de las investigaciones científicas, nos habla sobre la incorporación de la digitalidad en la vida social global. Individuos, personas, entidades, servicios, empresas y gobiernos, han incorporado progresivamente el trabajo digital y los procesos automáticos de procesamiento de información, de agendamiento de tareas, de observación de indicadores, de construcción del proceso de visualización de los datos, para citar algunos ejemplos, en sus vidas personales y laborales. Lo mismo pasa con la ciencia, fruto de su inserción en el mundo, un tiempo y un contexto tecnológico.

Me parece que la reflexión entre el hacer científico y los saberes producidos en las investigaciones contemporáneas en las Ciencias Sociales latinoamericanas pueden aportar de manera significativa al entendimiento de la ciencia que se hace en la actualidad. Sobre todo si tratamos de interpretar los estudios científicos como evidencias de una vida que pasa afuera de los laboratorios, en la sociedad. El hacer científico está evidentemente inscrito en un tiempo y contexto social. La introducción de los procesamientos computacionales de los grandes datos en la ciencia, es una de las muchas demostraciones de este proceso.

Debido a todas las implicaciones relativas a la naturaleza de los macrodatos discutidas anteriormente, se entiende que estos deben ser estudiados bajo una perspectiva sociológica crítica, ya que son un ensamblaje de artefactos tecnológicos (herramienta para el análisis de lo social), a la vez que se constituyen como un indicador, un espejo útil para profundizar en la comprensión de los comportamientos humanos (Nascimento, 2016).

Este trabajo se inscribe en el campo de las CTS<sup>20</sup> pues cree que la ciencia, tanto sus prácticas como contextos, influyen y se afectan por los aspectos sociales de la vida humana. En específico, a partir de las lentes, miradas y acercamientos metodológicos provenientes de la corriente constructivista de los estudios de la ciencia, tecnología y sociedad difundidos por Knorr-Cetina (1981), Michael Mulkay (1979), Michel Callon (1986), Lynch, M. (1983) y Bruno Latour (1995, 2005, 2011).

Se considera, entonces, que los estudios de la ciencia y el enfoque de la CTS son centrales para entender la problemática planteada en este proyecto de investigación. Se opta por entender la construcción social de conocimiento en apoyo del abordaje teórico desarrollado por los científicos sociales de la segunda generación del Programa Fuerte<sup>21</sup> (Science Studies Unit) de la Facultad de Ciencias de la Universidad de Edimburgo.

Esta escuela socio-antropológica fue la responsable de introducir en las ciencias sociales una nueva manera de investigar los fenómenos científicos. La sociología/antropología del conocimiento científico (SCC) producida por John Law, Michael Mulkay, Karina Knor-Cetina, Michel Callon y Bruno Latour. Se rescata de estos investigadores el planteamiento del concepto de *artefacto socio técnico* (Latour, 2011). Para ellos, la construcción del conocimiento científico está íntimamente vinculada a las características que el artefacto impone. Es el resultado de la interacción entre diversos elementos heterogéneos, efectos de múltiples operaciones realizadas por distintos representantes e intermediarios humanos y no humanos que se encuentran reunidos y articulados en una *red socio técnica* (Latour, 2011).

---

<sup>20</sup> Michael Mulkay (1979), Michel Callon (1986), Lynch, M. (1983), Karina Knor-Cetina (1981), Steve Woolgar (1979) y Bruno Latour (1995, 2005, 2011), obviamente.

<sup>21</sup> “El PF estableció cuatro principios programáticos. El principio de *causalidad*, que establece que las explicaciones que han de darse a estos procesos de producción tienen que ser de tipo causal –aunque bien entendido que no ha de concebirse la causalidad *sensu stricto*, al modo determinista o probabilista propio de las ciencias naturales, sino más bien en el sentido interpretativo y hermenéutico característico de las ciencias sociales (algo explica algo y, entonces es causa de ello, si le confiere sentido e inteligibilidad). El principio de *imparcialidad*, que establece que las razones que aduzca el investigador no pueden obviar las razones de los sujetos investigados; no hay que prejuzgar la definición de la situación que sus actores promulgan sino tenerla en cuenta como un elemento, y un elemento relevante, de la propia situación. El principio de *simetría*, en virtud del cual se postula que el mismo tipo de explicaciones que se den para los casos de conocimiento finalmente sancionado como exitosos o «buenos» han de ser aducidas para aquellos otros en los que de lo que se trate sea de intentos «fallidos», esto es, de procesos cuyo resultado, a posteriori, fueron juzgados como erróneos o malos. Por último, el principio de *reflexividad*, según el que, dado que la pretensión del PF es una explicación *científica* de los procesos de producción de conocimiento científico, las mismas explicaciones que desde él se produzcan, deben poder ser igualmente aplicadas sobre sus propios productos y afirmaciones.” (Ferreira, 2001, p. 109)

El contenido de este texto, también científico, fue construido a partir de las bases conceptuales con las cuales vengo estudiando a lo largo de mi investigación. La mirada apunta al objetivo de comprender big data no como objeto pasivo, sino como un artefacto, sujeto activo y participante del proceso de construcción de los saberes involucrados a su contexto. La aplicación de big data envuelve negociaciones, agenciamientos y principalmente traducciones entre todos los niveles de interacción entre los distintos actores (humanos o no humanos).

Parte del proceso de considerar un artefacto no humano, miembro activo de la construcción del conocimiento, significa sacar el foco del objeto, de su materialidad y limitaciones en cuanto técnica. Considerar un artefacto agente, en este trabajo, significó estar atenta a las relaciones e interacciones creadas a su alrededor en el proceso de la investigación. Big data, entonces, deja de ser una técnica, un amontonado de diversos datos, y se constituye como un agente influyente, imponiendo diversas e imperceptibles condiciones al entorno adyacente. Al trabajar los datos a través de big data, ciencia de datos, o procesamiento computacional de la información, la estructura de la investigación cambia juntos. Hay nuevos criterios de selecciones, otra disposición material y lógica de trabajo, que siendo la fuerza brazal de cualquier laboratorios de investigación, tiende a ajustarse conforme lo que complementa e interactúa con esta especie de motor de la investigación.

Tal cual en cualquier otro núcleo o laboratorio de investigación científica, el contexto de trabajo con/a través de big data, también hace interactuar prácticas científicas, científicos, sus técnicas de observación del mundo y los artefactos (no-humanos). Ellos se vuelven cómplices de unas acciones interrelacionadas, como bien relataba Latour (2008). Los no humanos, en su teoría, ganan entonces un papel de agencia activa, desempeñando, muchas veces, un rol diligente a la hora de valorar y sobresalir en determinadas controversias (Latour, 2008; 1995).

De igual manera en que ocurre en la vida social, Latour nos advierte sobre los ensamblajes entre humanos y no-humanos en el mundo científico:

“Cada disciplina ha elegido desplegar cierto tipo de mediadores específicos y ha privilegiado cierto tipo de estabilización, poblando así el mundo de diferentes tipos de habitantes, bien armados y enteramente formateados” (Latour, 2008, p. 372). Se busca, entonces, acceder a nuevas maneras de comprender las teorías, a partir de la formulación de nuevas preguntas, generadas por medio de la institución de prácticas investigativas alternativas, empezando por su propia investigación”. (Mendoza, 2001).

Entonces, es a partir de esta perspectiva que entendemos big data como uno más de los tantos artefactos no humanos que se relacionan, determinan y moldean el comportamiento humano, en este caso el científico. Se plantea que los macrodatos son un tipo peculiar de dato (digital) sobre la sociedad y contribuyen a la transformación de las prácticas investigativas del contexto en que está inserto. A su vez, se ha constituido como uno de los recursos tecnológicos contemporáneos cada vez más apropiados por la ciencia (Mendoza, 2001).

La construcción de un hecho científico, que también es un proceso persuasivo, culmina con un modelo de traducción que no se constituye sólo como la operación de comprensión y transmisión de saberes, sino que también en ella participan múltiples recursos e intereses que no tienen nada que ver con la “verdad” incorrupta a descubrir, sino con factores económicos, sociales, políticos, psicológicos, emocionales. Es decir, aspectos constitutivos de lo social (Latour, 2008).

Defiendo que la utilización masiva de técnicas de procesamiento de grande datos, como se ha explicado hace big data, es un buen ejemplo para pensar la íntima relación que hay entre método, técnica de producción de información, práctica científica y construcción del conocimiento científico y de la sociedad, en nuestra sociedad contemporánea en la era digital de la información.

En el contexto específico de esta investigación, como también he mencionado a lo largo del capítulo introductorio, encontré entre las diversas lecturas en los estudios de ciencia y tecnología, con los textos y la perspectiva teórica de Karin Knorr Cetina (2005). Me vi, poco tiempo después, sumergida en la lectura de *La fabricación del conocimiento. Un ensayo sobre el carácter constructivista y contextual de la ciencia*, edición de 2005, en español, editada por Pablo Kreimer, quien me concede la gentileza de ser miembro evaluador de este proyecto de tesis.

Este texto me tocó intelectualmente. Tal vez haya una identificación por la manera sensible como escribe, por el hecho que la autora también ser una mujer, o la forma como describe y prescribe su hacer etnográfico, su entrada al campo. El hecho es que esta obra me marcó positivamente, influenciando la manera en que observo, interactúo y analizo los hallazgos de mi propio terreno e investigación.

Knorr-Cetina es una socióloga y antropóloga austriaca, conocida por sus contribuciones en el campo de la sociología del conocimiento y la ciencia. Es especialmente reconocida por su teoría de la "epistemología de la práctica" que se centra en cómo los científicos adquieren conocimientos mediante la observación y la participación en prácticas científicas (Knorr-Cetina, 2005).

Centrándose en la comprensión de cómo la ciencia y la tecnología construyen la sociedad, al mismo tiempo que las utiliza como sustrato para la construcción de sus saberes, la autora ha hecho importantes contribuciones al campo de la etnografía de la ciencia, focalizando sus trabajos en cómo se produce el conocimiento científico en la práctica, a través de la observación directa y la participación en la actividad científica en sí misma. La perspectiva de Knorr-Cetina en la etnografía de la ciencia se ha enfocado en aspectos como la tecnología, la materialidad y la cultura científica, lo que la ha llevado a ser reconocida como una de las principales exponentes de la corriente constructivista en los estudios de ciencia, y mi gran mentora en este camino.

También conocida por su trabajo y difusión de la antropología sensitiva, la autora busca demostrar cómo la experiencia sensorial y corporal de los individuos y su relación con el mundo social y cultural en el que viven impactan en sus decisiones cotidianas y culturales. Se basa en la idea de que nuestras experiencias sensoriales no son simplemente "datos brutos" que se procesan en el cerebro, sino que están mediadas por nuestras relaciones sociales y culturales. Knorr-Cetina defiende que la ciencia y los estudios de la ciencia, igual que cualquier otra dimensión de la vida social, es impactada y construida de acuerdo a las experiencias corporales y sensoriales que uno vivencia en el terreno de la investigación (Knorr-Cetina, 2005). Según ella, los científicos adquieren conocimientos a través de la experiencia sensorial y la participación en prácticas científicas. Ya no es suficiente apenas observar, hay que involucrarse con las diferentes dimensiones de la investigación. Estar de cuerpo presente.

Esto era lo que justamente pasaba conmigo durante la investigación. Ya no era solamente una observadora en el laboratorio que me propuse estudiar. Me transformé en una parte involucrada en las actividades de este núcleo: abarcaba mis esfuerzos, cerebro, pensamiento, opiniones, y muchas veces, ¡muchos dedos en la computadora! Hasta volverme yo misma una trabajadora más del laboratorio que investigaba. Mi propia jornada laboral, pasa entonces a ser objeto de estudio y atención.

Es por esto que la perspectiva de Knorr-Cetina, 2005, se consolida como una entrada teórica muy útil a este trabajo. La perspectiva de la autora se consolida aquí como un enfoque teórico esencial para este trabajo, sirviendo tanto de marco teórico principal como de modelo metodológico de análisis. A pesar de reconocer el valor de otros autores constructivistas de los estudios de ciencia, tecnología y sociedad (CTS), opté conscientemente por centrarme casi exclusivamente en los modelos teórico-metodológicos de Knorr-Cetina. Esto se debe a que considero que su obra es conceptualmente, teóricamente y metodológicamente suficiente para abarcar los aspectos investigativos dentro del alcance de este estudio. Como señala Marcel Mauss (2003) en su célebre obra *Ensayo sobre el don*, al explorar el concepto de "hecho social total," defiende la libertad de elegir una obra que, en su totalidad, pueda responder de manera integral a las problemáticas planteadas en esta investigación. En suma, esta obra, fue un importante guía conductor de mi mirada al analizar las experiencias que viví durante la etapa etnográfica de esta investigación, y también responsable por conducirme a la reflexión sobre mi propia intervención en el ambiente que investigo.

Su entendimiento de que el científico como sujeto condiciona la experiencia del campo, a la vez que está condicionado por las peculiaridades y contextos en donde está insertado, me dirigió al entendimiento y a la responsabilidad de posicionarme con respecto a las especificidades que se hacen parte de mi propio proceso investigativo: ser latinoamericana, mujer y migrante, en un contexto de investigación perteneciente a países del sur global, en una universidad pública. Su lectura es en gran parte responsable de mi perspectiva etnográfica declaradamente situada, discutida con más detalles en los apartados que se siguen.

## **2.5 La investigación social latinoamericana**

El desarrollo de la ciencia está íntimamente ligado a los contextos de investigación manifiestos a su alrededor. Esto también quiere decir, según lo que entiendo, estar atento a las condiciones que pertenecen, o no pertenecen, al ambiente propicio para el desarrollo de diversas exigencias, las que también son parte del proceso de investigar de forma situada y constructivista.

Para manipular macrodatos, o “trabajar *bigdatamente*”, es necesario una estructura mínima de operación, capacidades y letramento. Como consecuencia, los centros de investigación,

núcleos de pesquisa, observatorios y demás iniciativas científicas dispersas por todo el mundo, están estructuradas de acuerdo a la disponibilidad de recursos e infraestructura operacional y humana que logran garantizar para sí.

Emprender investigaciones científicas nunca fue más fácil en países del globo sur. Los deconolialistas (Mignolo, 2008; Quijano, 2014; Domingues, 2015) discuten brillantemente sobre las implicaciones de los distintos privilegios del norte global en todo el mundo. De igual manera, investigar bajo la lógica y estructura que impone big data implica también superar barreras tecnológicas, territoriales, económicas y legales, por citar apenas algunas áreas de desafío, que las instituciones científicas provenientes de contextos de más desarrollo no poseen de antemano.

Limitaciones de esta naturaleza dejan en desventaja a los laboratorios y científicos de la parte sur del globo. Sumado a este factor, está el hecho que en las investigaciones sobre big data, el centro o el “motor” de las investigaciones, para citar palabras nativas, *es* el dato. Tener la posibilidad de almacenamiento y estructura de operación adecuada, humana y no humana, para procesar y analizar las informaciones y los datos disponibles es una ventaja que apenas pocas empresas, norteamericanas poseen (Gurstein, 2011).

Es por esto que como científica latinoamericana, la construcción de una investigación situada en el contexto que vivo y la territorialidad que experimento mientras desarrollo mi investigación, es decisiva. Este es un trabajo interesado en comprender los aspectos investigativos pertinentes a una ciencia latinoamericana y que evidencie el rol de las ciencias sociales en el contexto actual del continente, identificando los desafíos existentes para la generación de nuevas teorías sociales a partir de problemáticas contemporáneas y locales.

Mi interés por las investigaciones latinoamericanas en big data está en la búsqueda por responder lo que Domingues (2015) defiende: la necesidad de contar con una nueva teoría de la modernidad global, que se sustente en una perspectiva emancipadora-crítica. Que recoja la herencia teórica europea y estadounidense, pero, al mismo tiempo, incorpore las perspectivas latinoamericanas y de otros países periféricos.

De este modo, defiende la necesidad de entender el fenómeno de los macrodatos de manera global (las potencialidades para el desarrollo de nuevas teorías sociales donde sea que opere) y a la vez, de forma particularizada. Para emprender dicha tarea y construir una epistemología

que sustente la investigación y el desarrollo teórico en nuestro continente latino, cabe preguntarse si sería necesario abrir una reflexión en torno a ‘lo social’ global, sin tampoco reducir los análisis e interpretaciones como resultado de sus peculiaridades locales.

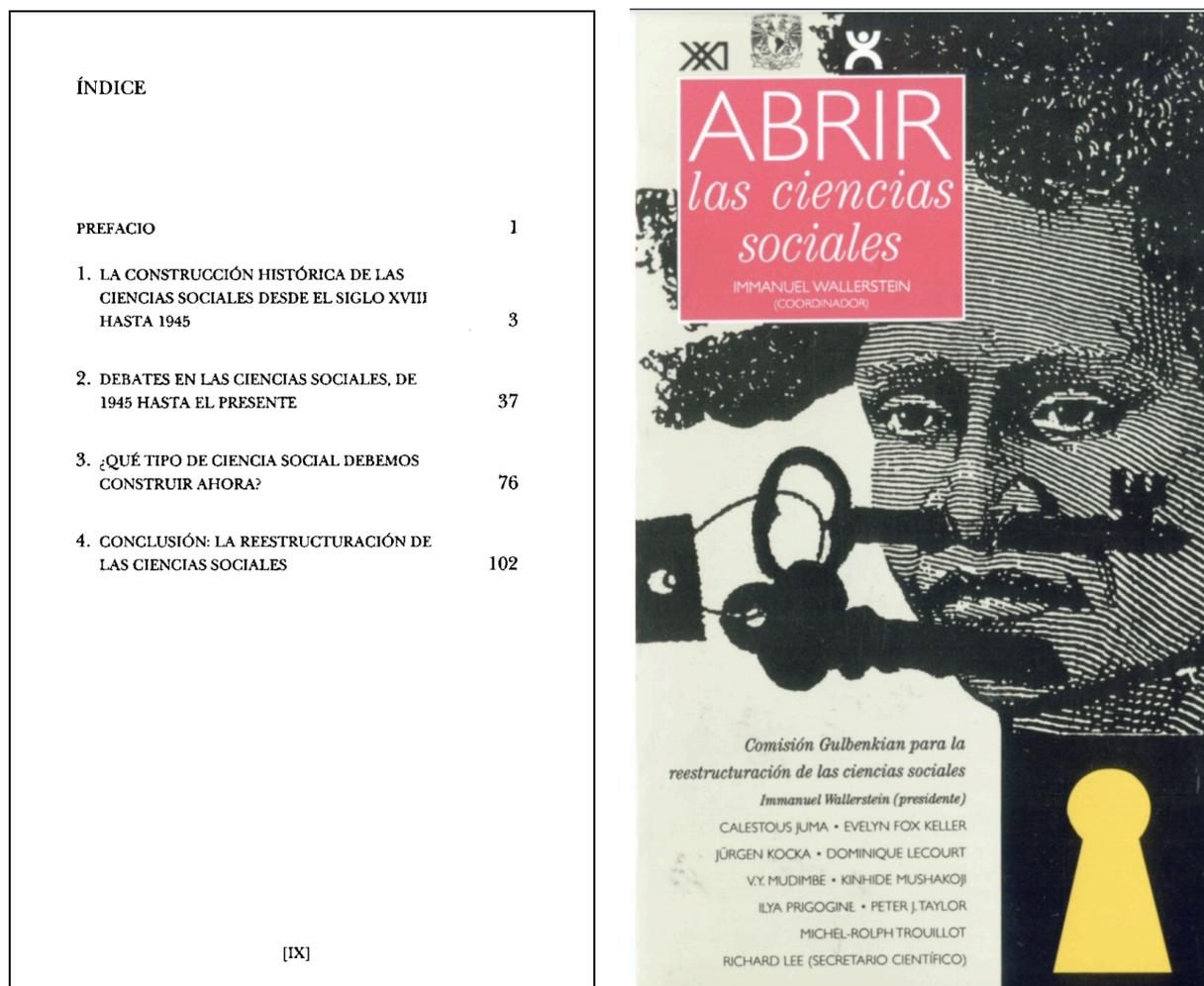
Los estudios revisados en la primera fase de esta investigación relevan – directa o indirectamente – la importancia del rol de las ciencias sociales para aportar a la reflexión en torno a las problemáticas actuales latinoamericanas en el contexto global. Hay un movimiento creciente de especialistas sociales interesados en la introducción de métodos computacionales que necesitan ser evidenciados: desarrollan soluciones y sistemas innovadores, creados a base de mucho amor por el oficio y ganas de mejorar sus actividades y resultados, aunque no tengan los medios apropiados para ello. Como suele pasar en cualquier laboratorio científico latinoamericano.

Es por eso que los objetivos de esta investigación están dirigidos a la comprensión de la relación entre técnica, práctica, contexto científico, la producción y reproducción de las narrativas sobre nuestras sociedades y que vienen siendo construidas al interior de las ciencias sociales latinoamericanas a través del big data. Es, también, un intento de generar el espacio para una reflexión teórica de carácter interdisciplinario, de modo que el desarrollo de nuevas teorías localmente situadas sean capaces de aportar a la reflexión de la complejidad social actual, cada vez más global (Quijano, 2014).

Con respecto a la producción de conocimiento científico por las ciencias sociales, entendemos que, sus acciones se han dirigido históricamente a dos objetivos principales: el estudio de ‘lo social’, que ha buscado comprender qué es lo propiamente social, estableciendo regularidades; y el estudio de ‘las sociedades históricas’, donde se han analizado sociedades particulares. Lo anterior cobra suma relevancia si se considera que, desde la perspectiva de diversos autores, las ciencias sociales han dado respuestas “anacrónicas, desajustadas y – en el mejor de los casos – incompletas” a diversos problemas sociales (Follari, 2015, p.1).

La Comisión Gulbenkian<sup>22</sup> fue otra iniciativa llevada a cabo para plantear la inquietud de las ciencias sociales en torno a su insuficiencia epistemológica para comprender las sociedades contemporáneas: así desde su instalación, las ciencias sociales apuntaron más bien a comprender la naturaleza específica de las sociedades históricas (Garretón, 2000).

**Figura 6.** Portada e índice de *Abrir las Ciencias Sociales*, 1995.



A la fecha, las ciencias sociales latinoamericanas se han caracterizado por una marcada carencia de teorización, incorporando nuevamente – con fuerza y a veces de manera acrítica – teorías europeas y estadounidenses (Domingues, 2015). Al respecto, es importante destacar el

<sup>22</sup> La Comisión Gulbenkian buscó superar las insuficiencias en la organización de las disciplinas de las ciencias sociales que se desarrollaron en el siglo XIX, indicando una dirección para la investigación en ciencias sociales para los próximos 50 años. Fue fundada por la Fundación Calouste Gulbenkian. Celebró tres reuniones en 1994 y 1995. Para saber más: <http://paginaspersonales.unam.mx/app/webroot/files/5880/Asignaturas/1876/Archivo2.4541.pdf> accedido en 11 de mayo de 2023.

esfuerzo que, desde principios de los noventa, pero con raíces anteriores, han desarrollado grupos de investigadores bajo el pensamiento poscolonial (Follari, 2015) más específicamente, decolonial (Mignolo, 2008).

Cabe señalar, sin embargo, que la existencia de estos grupos intelectuales caracterizados como decoloniales no ha logrado contrarrestar el peso de las teorías importadas desde el norte (Follari, 2015). Junto con lo anterior, el pensamiento decolonial habría plasmado una mirada reductiva de la modernidad, centrándose únicamente en sus características opresivas, sin considerar su dimensión emancipatoria. Además, sus aportes serían débiles en cuanto a rigor científico, basándose a veces en una retórica especulativa (Domingues, 2015).

Sumado a esto, está el intercambio internacional que se produce desde la construcción de los saberes –para América Latina– en el marco de una dependencia tecnológica, comunicacional, financiera y de expresión cultural (Castells & Calderón, 2004). El reconocido debilitamiento de la cultura científica social latinoamericana se justifica en base a la carencia de recursos y de una masa crítica de investigadores en la región, lo que implica –a su vez– serias dificultades para abordar la brecha que hay frente a los países ‘desarrollados’ (Garretón, 2000).

Este patrón también es percibido en la problemática del presente trabajo. Percibimos una significativa desproporcionalidad de investigaciones, grupos de estudio y laboratorios de investigación que se ocupan de la temática big data en Ciencias Sociales. Mientras que según la base de datos Scopus, un 46% de las publicaciones que responden a investigación social en big data son provenientes de China y de Estados Unidos, y apenas el 1,4% del total de 7760 artículos son escritos por autores latinoamericanos. De los 33 países pertenecientes a América Latina y el Caribe, apenas 7 poseen algún tipo de publicación en Ciencias Sociales y big data. Este número es aún menor si hacemos la misma búsqueda en la base de datos de Scielo: de los 255 resultados “big data and social science”, apenas 9 son provenientes de América Latina<sup>23</sup>.

---

<sup>23</sup> Investigación levantada a través de la página web de los citados bancos de datos:

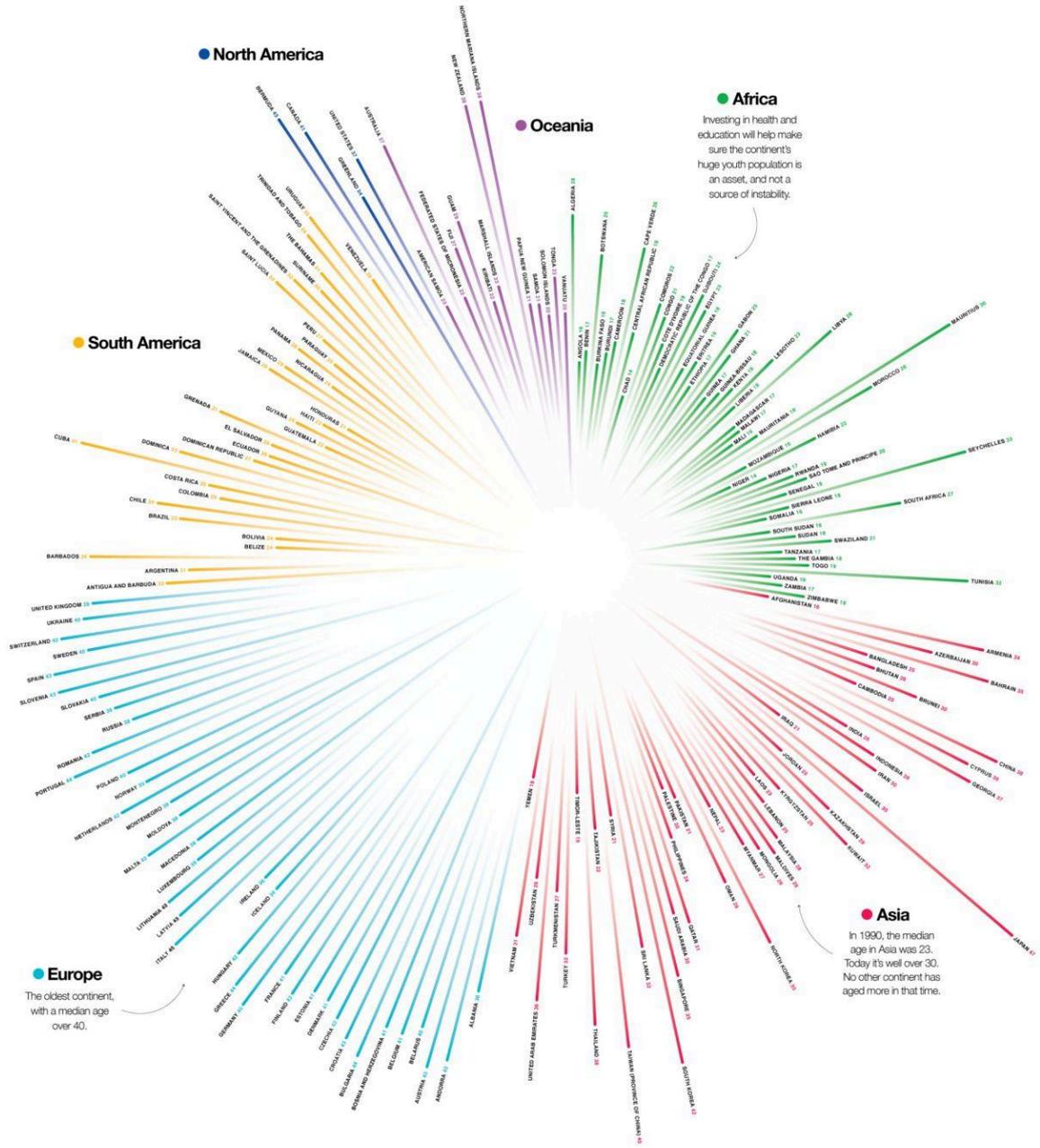
a) <https://www.scopus.com/sources?zone=TopNavBar&origin=NO%20ORIGIN%20DEFINED> y b) <https://www.scielo.org>.

Esta misma discusión pasa por un segundo dilema: la concentración de conocimiento para acceder a los datos, procesarlos y analizarlos resulta también un desafío para las ciencias sociales latinoamericanas interesadas en los grandes datos. Este no es un argumento menor, ya que la falta de datos y de conocimiento especializado se encuentran focalizados en ciertas universidades estadounidenses y europeas, en comparación a la discreta producción de regiones como América Latina (Gurstein, 2011).

Paradójicamente, en el caso de las redes sociales digitales, América Latina es la región que más las utiliza en el mundo (Cepal, 2015:63). La Cepal señala que es en esta región donde se encuentra el uso más intenso de las redes sociales, con una participación de 78.4% de los usuarios de Internet en 2013. El organismo atribuye este fenómeno a los bajos costos de hardware y a la globalización de patrones de consumo, así como a la débil oferta latinoamericana de redes sociales digitales y de buscadores. Esta participación en la generación de datos, paradójicamente, no se traduce en más investigación regional. El algoritmo de Google Académico encuentra 75900 textos sobre big data del año 2000 al año 2016, en contraste con 3510 en español en el mismo lapso. La región latinoamericana es una gran fuente de datos para las empresas de la economía digital, que los blinda en sus cajas negras para reformular sus negocios (Rocha, 2018, p. 427).

Es por eso que se considera fundamental emprender una investigación que evoque la discusión entre condiciones de producción del conocimiento científico social respecto a su propio contexto de producción. Aún más: Latinoamérica, como la región del mundo que más utiliza las principales fuentes de recolección de datos socio-digitales, sigue siendo el territorio que menos produce saberes asociados a la problemática de la digitalidad de los datos. Se espera poder entender y contribuir, entonces, a partir del privilegio de análisis del contexto de producción científica de big data en Latinoamérica, con la consolidación del desarrollo de los estudios de Ciencia, Tecnología y Sociedad sobre el fenómeno en Latino América.

# PARTE 2: TEXTO



“La escritura del artículo comienza mucho antes de que esté escrito el borrador, mediante las huellas que se han ido generando durante el trabajo *de laboratorio*” (Knorr-Cetina, 2005, p.285).

### Cap 3: Scielo, una librería científica latinoamericana.

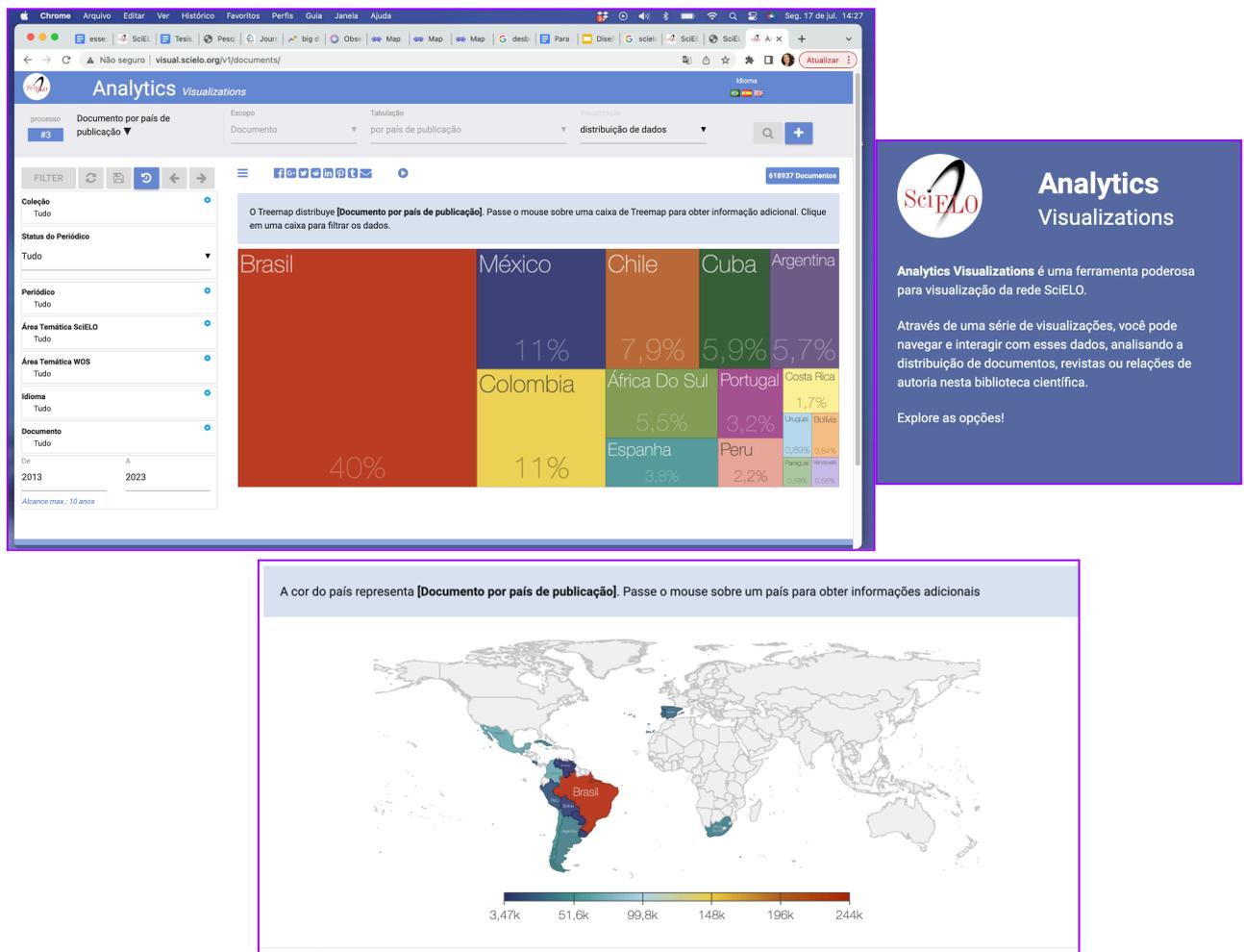
Para iniciar la investigación bibliométrica, se utilizó la base de datos Scielo, ampliamente empleada y accesible en Brasil. Además, en investigaciones previas para la construcción del marco teórico (descritas en el capítulo 2), se observó que una parte significativa de las publicaciones latinoamericanas en ciencias sociales y big data procedían de autores o revistas brasileñas. En la fase de revisión bibliográfica, de los 45 textos seleccionados en ciencias sociales, 15 provenían de revistas latinoamericanas. De estos 15, 9 (60%) estaban escritos en portugués, y otros 3 provenían de revistas brasileñas, pero estaban en español. Los textos restantes incluían 1 artículo en español con autores brasileños y 2 artículos en español de revistas hispanohablantes.

Esta elección inicial resultó además estratégica. Las imágenes a continuación<sup>24</sup>, obtenidas de la sección de “visualización analítica” del portal SciELO, corroboran esta tendencia. Brasil es el país con el mayor número de documentos en SciELO, representando el 40% del total (240 mil publicaciones en todas las áreas temáticas). México y Colombia ocupan el segundo lugar, con 66,900 y 65,900 textos respectivamente, lo que equivale al 11% de las publicaciones en la base.

---

<sup>24</sup>Este es el enlace para acceder el Analytics de Scielo. <http://visual.scielo.org/v1/>. Lo estoy accediendo en el día 17 de julio de 2023, al momento de hacer la primera corrección de este apartado, después de casi 2 años de escrito por primera vez. Al revisarlo con objetivo de destinar el texto a la persona que contraté para que me ayudara con las correcciones de ortografía y formación del texto, decido que esta información debería ser corroborada de manera más científica. Sentí que faltaba pruebas que soportan mi decisión en empezar la búsqueda por Scielo. Resuelvo acceder una vez más al portal de Scielo buscando más informaciones sobre la base de datos, y en un par de clics, me deparo con esta nueva aba en su portal. Cosa que no existía en el año de 2021, o sea, 2 años atrás. El desarrollo y acceso de los recursos de visualización de datos (interactivos) en la ciencia y su accesibilidad son impactantes. El propio Scielo lo divulga con entusiasmo y un cierto tono de enaltecimiento: “Analytics visualization es una herramienta poderosa para la visualización de la red SciELO” [mi traducción], como se puede apreciar en la figura 7: Página de Analytics Scielo accedida en 17 de julio de 2023.

Figura 7. Publicaciones por país de Scielo entre 2013 y 2023



Nota. Captura de pantalla de Scielo.com

Como se puede notar la presencia de países latinoamericanos en Scielo es masiva. Por lo menos 83% de las publicaciones indexadas a su base de datos es latinoamericana. Por ello, resulta fundamental para esta investigación.

Otro factor decisivo que justifica la incorporación de la base de datos Scielo a los estudios bibliométricos de la presente investigación es la probabilidad de abarcar una mayor cantidad de publicaciones escritas en portugués y español, condición que se considera esencial a esta investigación, dado que el enfoque es analizar los estudios científicos sociales y latinoamericanos.

La alta aceptación de idiomas latinos, facilita mucho la inserción y difusión de trabajos de científicos latinoamericanos. En las “Ciencias humanas”, el idioma en que más se publica es

el español, con un total de 277 mil textos, seguidos del inglés, 264 mil, y el portugués con 142 mil publicaciones. Juntos, español y portugués son el idioma predominante en esta base de datos (más de 419 mil documentos), lo que corresponde a un 61% de todas las publicaciones en Scielo entre los años de 2013 a 2023.

Por último, Scielo posee una gran cantidad de revistas en humanidades y ciencias sociales. Según *Analytics Visualization* de Scielo, la mayoría de las publicaciones vinculadas a su base de datos son de las “ciencias de la salud” con 242 mil textos. Pero las áreas “Ciencias Humanas” (131 mil textos) y “ciencias sociales aplicadas” (102 mil), juntas, poseen un número de publicaciones (233 mil) muy próximas a las disciplinas de salud. Sea como sea, es innegable que humanidades y ciencias sociales también poseen un gran número de publicaciones indexadas a librería, por lo que es igual de importante de examinar, cómo podemos verificar en la figura siguiente:

**Figura 8.** *Scielo Analytics: Proporción x área temática*



*Nota.* Captura de pantalla de Scielo.com

### 3.1 Parámetros de búsqueda: como se hizo para obtener tales resultados

El 1 de febrero de 2021 comienza el proceso de análisis de la literatura sobre big data en publicaciones de ciencias sociales en Scielo. Al acceder a la página principal de “Scientific Electronic Library Online” se inicia una búsqueda precisa mediante la introducción del término “big data” entre comillas para afinar los resultados. Esta búsqueda inicial arroja 15

páginas con un total de 223 textos recomendados, en donde, entre los primeros 15 resultados, solo 2 contienen la palabra "big data" en el título.

Tras un breve periodo de exploración preliminar, se procede a utilizar la función de "búsqueda avanzada" con el fin de filtrar específicamente los textos de ciencias sociales relacionados con big data. En el menú de filtros, ubicado a la izquierda de la página, se seleccionan las opciones de país, idioma, año de publicación, áreas temáticas de SciELO, tipo de literatura, entre otros. Al mantener el término "big data" como criterio de búsqueda principal y aplicar el filtro en las áreas temáticas "ciencias humanas", el sistema reduce los resultados a 26 textos, representando un 11,6% del total inicial de textos sin filtros temáticos (ver tabla 1).

**Tabla 1.** Comparación entre búsqueda de "big data" en todas las áreas temáticas vs. ciencias humanas.

Término de búsqueda	Áreas temáticas	Número de textos	Porcentaje del total
"big data"	Todas	223	100%
"big data"	Ciencias Humanas	26	11,6%

De los 26 textos encontrados, 15 de ellos provenían de Brasil, seguidos de Colombia, México, Paraguay, Chile y Cuba (tabla 2). Observe como las publicaciones brasileñas representan más de la mitad de los textos que posee big data como término incluido en el motor de búsqueda de Scielo. Este resultado era lo que uno ya podría esperar si consideramos que Brasil se constituía como el país con mayor participación en la publicación de textos.

**Tabla 2.** Publicaciones sobre big data por país en Scielo.

Países	Nº literatura	Porcentaje del total
Brasil	15	57,6%
Colombia	5	19,2%
México	2	7,6%

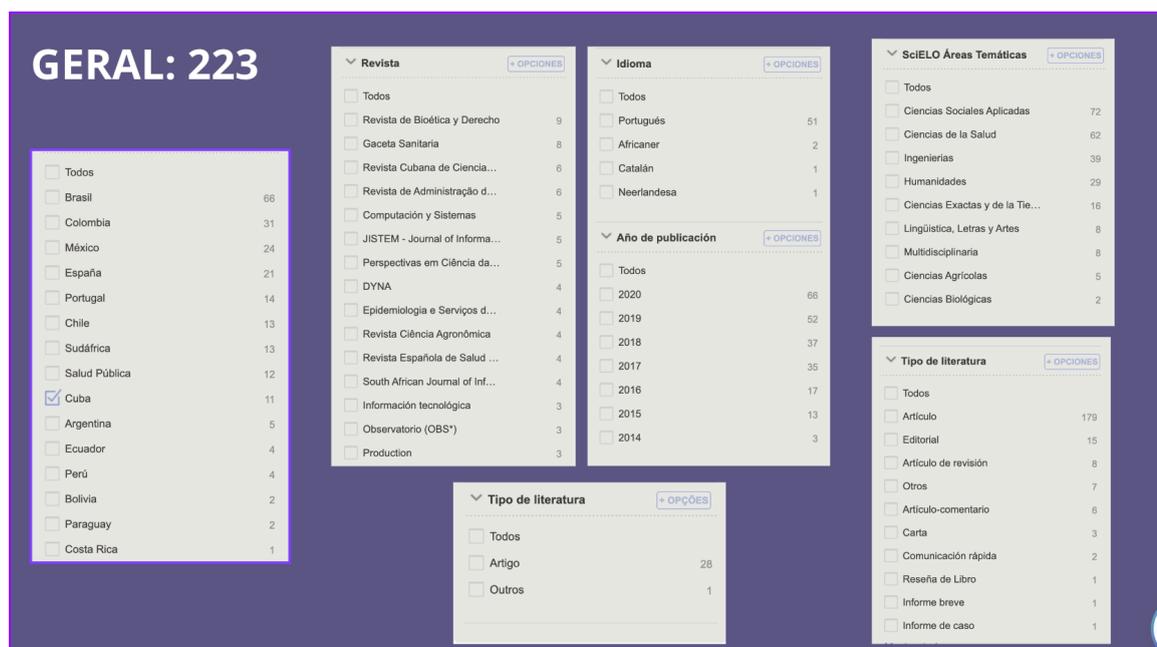
Paraguay	2	7,6%
Chile	1	3,8%
Cuba	1	3,8%

Con respecto al tipo de literatura asociado al término big data en las ciencias humanas en la base de datos Scielo, hasta diciembre de 2020, el artículo sigue siendo el tipo de texto más publicado. 25 textos, contra apenas 1 en “otros”.

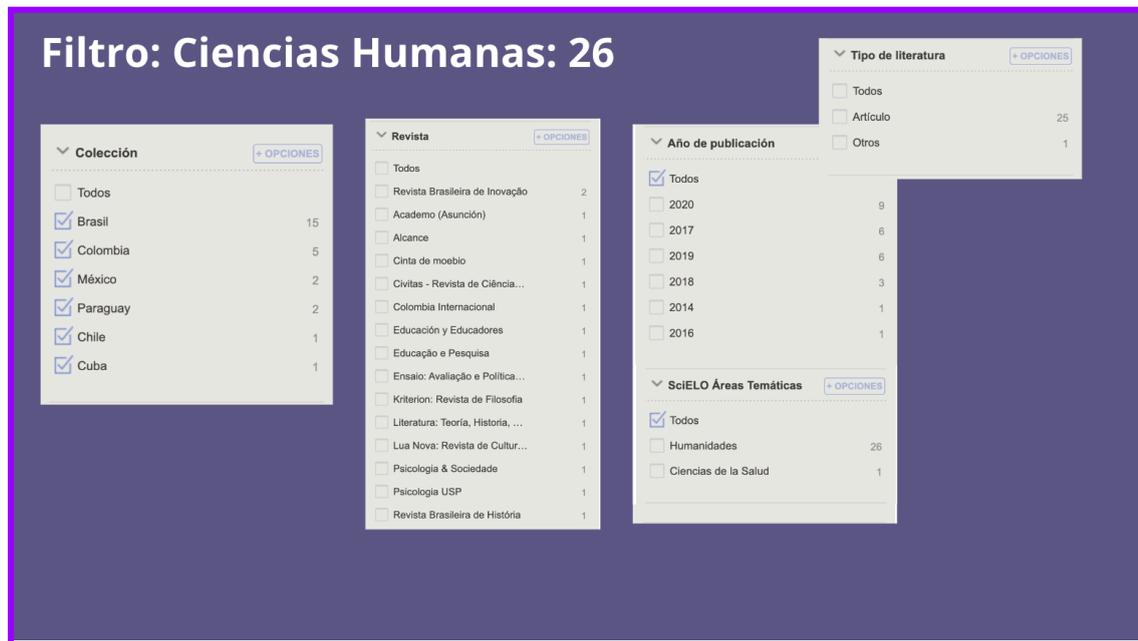
En cuanto a los periódicos que más presentan publicaciones según estos mismos criterios de búsqueda son: *Revista Brasileira de Inovação* con 2 artículos, seguidas por todas las otras, con 1 texto cada una.

En la secuencia, les presento dos figuras comparativas. La primera (figura 9) corresponde a los resultados de la búsqueda por publicaciones científicas sobre big data en todas las áreas temáticas, según los índices de la propia librería Scielo. La figura 10 corresponde a las publicaciones de ciencia sociales.

**Figura 9.** Base Scielo: “big data” + todas las áreas temáticas



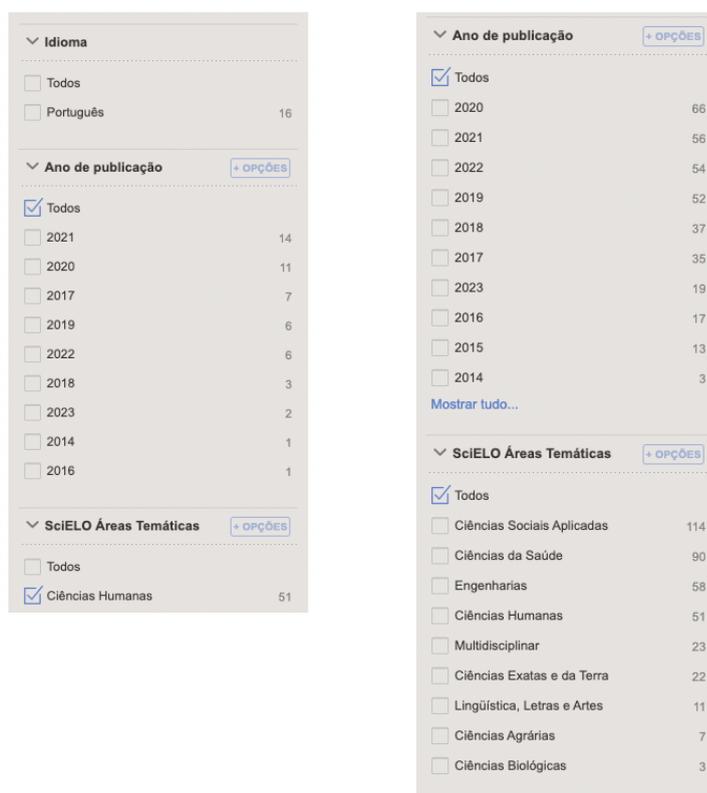
**Figura 10.** Base Scielo: búsqueda de “big data” + “ciencias humanas”



Para ambas búsquedas, 2020 fue el año de mayor número de publicaciones sobre big data, al contrario del momento de escritura de este capítulo. A la fecha, de los 26 textos publicados hasta diciembre de 2020, saltamos a un total de 51 publicaciones en Ciencias Humanas, hasta el primer semestre de 2023 (figura 11).

La progresión de publicaciones año por año puede verse en la siguiente figura. A la derecha, tenemos los textos referentes a las “ciencias humanas”, y a la izquierda, el número de publicaciones por año de “todas las áreas temáticas”.

**Figura 11.** Datos bibliométricos “CS” x “todas las áreas”, hasta 2023



Se observaron un total de 359 publicaciones en todas las áreas temáticas, hasta la primera mitad de 2023, contra 223 textos publicados hasta febrero de 2021, representa un crecimiento acentuado en poco más de 2 años. 2020 es el año líder en publicaciones, 66, hasta el final de este año. Después está 2021 con 56 textos. En tercera posición, el año 2022 con 54 textos. 2023 ocupa el 7º lugar, con un total de 19 publicaciones, por lo menos hasta la fecha, mayo 2023.

En ambas áreas temáticas, Brasil y Colombia siguen liderando el ranking de los países con mayor número de textos indexados a esta base de datos tanto en 2021, como hasta la actualidad (2023). Con relación específicamente a las disciplinas que más publicaron en Scielo textos relacionados a big data hasta 2023, están “ciencias sociales aplicadas” con 114 publicaciones, seguida de “ciencias de la salud” (90), y en tercer lugar “ingenierías” (58).

Es relevante señalar que Chile tiene una representación inusual en relación con su volumen de publicaciones en esta temática. En el área de “ciencias humanas”, por ejemplo, cuenta con

solo una publicación<sup>25</sup>, pero de gran relevancia, acumulando 9.525 descargas en la biblioteca Scielo hasta la fecha, tres veces más que los tres textos principales de Brasil combinados. Esto podría atribuirse, en parte, a su título atractivo para investigadores en ciencias sociales interesados en el tema: “Big data, algoritmos y política: las ciencias sociales en la era de las redes digitales”.

Ya en Brasil, mayor representante de artículos científicos sobre big data en Scielo, hasta el año de 2023 posee el área temática “ciencias sociales aplicadas” en primera posición con 34 textos. Seguidos de “ciencias de la salud” (32); “ciencias humanas” (21) y “ciencias agrarias” (7).

El posicionamiento de las “ciencias humanas” en tercer lugar, según se observa, no indica una inserción masiva de estudios académicos en ciencias sociales sobre el tema. Esto se respalda en la diferencia del 13% de publicaciones en “ciencias humanas” en comparación con otras áreas. Este resultado sugiere que, en realidad, SciELO es una base de datos ampliamente utilizada en ciencias humanas, especialmente en Brasil.

La revista con más artículos publicados en Brasil hasta el primer semestre de 2023 es *JISTEM - Journal of Information Systems and Technology Management*<sup>26</sup>, con un total de 7 artículos. Esta revista, originada en la Universidad de São Paulo y mantenida por una asociación dedicada a sistemas e información, publica exclusivamente en inglés. En segundo lugar, con 6 artículos cada una, se encuentran las revistas *Revista de Administração de Empresas* y *Perspectivas da Ciência da Informação*.

Con el objetivo de obtener una comprensión general de los contenidos y los temas abordados en los textos, se realizó una revisión de los títulos de los artículos. En este contexto, se consideró la implementación de una técnica de visualización a través de nubes de palabras, que permite representar los términos presentes en los títulos de los 26 textos sobre big data en las publicaciones de ciencias humanas de Scielo. Esta técnica facilita la identificación de los términos de mayor peso y frecuencia, proporcionando una representación visual clara de las palabras asociadas al tema.

---

<sup>25</sup> González, Felipe. (2019). Big data, algoritmos y política: las ciencias sociales en la era de las redes digitales. *Cinta de moebio*, (65), 267-280. <https://dx.doi.org/10.4067/s0717-554x2019000200267>

<sup>26</sup><https://www.tecsi.org/jistem/index.php/jistem>: “JISTEM USP is the official Journal of the AIS Brazilian Chapter. (AIS is the Association for Information Systems) JISTEM USP is assessed by the Brazilian System Qualis/CAPEs”. Accedido en julio de 2023.

Herramientas como Wordle®, WordCloud® y plataformas como Python<sup>27</sup> permiten la generación de mapas conceptuales que visualizan los conceptos más relevantes en un conjunto de textos. Estos softwares ordenan y presentan infografías con los términos más frecuentes y su relación con palabras clave, lo que contribuye a un análisis más profundo de la relevancia y frecuencia de cada término en los documentos analizados. Esta estrategia de análisis se presenta como un recurso práctico y visual para identificar los contenidos recurrentes y mapear las relaciones entre diferentes términos, lo que puede abrir nuevas vías de descubrimiento sobre la frecuencia e importancia de los contenidos textuales.

Así, se utilizó estas herramientas para analizar los títulos de los 26 publicaciones sobre big data en la base de datos latinoamericana de Scielo. El objetivo no fue solamente construir un mapa bibliométrico de la producción científica en torno a big data, sino también ofrecer un panorama general de los contenidos presentes en dichas publicaciones. Al identificar los temas más frecuentes, las palabras clave más utilizadas y los contenidos asociados a temáticas específicas, contenidas en los títulos de dichas publicaciones, se espera encontrar pistas que faciliten la comprensión de las narrativas construidas en torno al conjunto de textos analizados<sup>28</sup>. La utilización de esta técnica complementa y enriquece el análisis de los contenidos de un gran número de documentos, más allá de los índices y frecuencias normalmente asociados al estudio bibliométrico.

La figura abajo muestra el resultado del conteo de palabras proveniente de los títulos de las 26 publicaciones en “big data” + “ciencias humanas” + “países latinoamericanos”. Las palabras que más se repiten, o sea, las con mayor frecuencia, corresponden a las palabras de mayor tamaño en la figura. Segundo aspecto a observar: cuanto más próxima una palabra de menor tamaño está de las principales, hay más chance de estar asociado a este término. Por ejemplo: en esta base de datos que estamos analizando, es más probable que el término big data esté asociado a “derechos” y “sociales”, pero también a “trabajo”, “humanos”, “equipos”

---

<sup>27</sup> Todos los programas son de libre acceso para cualquier usuario.

<sup>28</sup> Los softwares de nubes de palabras también permiten el análisis de frases cortas, de cuatro a cinco palabras (palabras claves de *long tail*) o la LSI (término en inglés para Indexación Demántica Latente) donde es posible sumar? a la búsqueda de las palabras más frecuentes y sus términos relacionados, en un intento de entender los conceptos que rodean el contexto de ciertas palabras claves (Prais y Rosa, 2017).

y “análisis”. La palabra “datos”, está fuertemente asociada a “inteligencia”, “artificial” y “protección”.

**Figura 12.** Nube de las palabras claves de las publicaciones de CS en Scielo.



Con el objetivo de comprender la relación entre los términos y las investigaciones producidas, se optó por descargar manualmente cada texto. Los archivos fueron organizados en una carpeta en el computador titulada "Publis Scielo: BD + CH + LA", lo que facilitó la posterior lectura detallada de cada documento.

Una vez completada la descarga, se inició un proceso de categorización de la información contenida en las publicaciones. Paralelamente a la lectura bibliométrica, se construyó una nueva base de datos utilizando una tabla de Excel. Cada vez que se descargaba una publicación, se registraban de inmediato los datos relevantes en la tabla, incluyendo las variables consideradas importantes: "título", "palabras clave", "referencias", "autores", así como categorías como "autor único", "dos autores", "más de tres autores", "revista de publicación", entre otras.

Durante este proceso, surgieron nuevas inquietudes y preguntas que llevaron a la incorporación de columnas adicionales en la tabla. La figura 13 que se presenta a continuación, ilustra las categorías creadas en esta nueva base de datos.

**Figura 13. Categorías/variables de la base de datos oficial de Scielo**

link	País		P	Año		Área	Nº	Vínculo						Nº	Nº	Nº	Nº
pub	de	T	al	P	R	discipli	A	I	País	I	Sexo	Canti	T	d	Nº	Nº	Nº
lic	la	it	bra	ubl	evi	nar	u	n	del	di	A	dad	i	e	big	big	big
ca	Re	ulos	as	ic	sta	de la	o	st	A	o	u	ter	p	s	data	data	data
ci	vis	C	Cl	ca	ca	Re	r	it	o	ma	t	mi	o	c	T	R	P
ón	ta	l	ave	ión	ta	v	e	u	o	del	nos	big	o	a	i	e	a
		s	s	n	a	s	s	o	r	texto	big	data	o	r	g	u	v
								n			data			a	l	l	r
								a						s	o	o	a
																	s
																	Claves

Con la creación de esta nueva base de datos y teniendo como base las categorías mostradas en la figura anterior, se obtuvieron los resultados que se presentan a continuación y que pueden consultarse en su totalidad a través del siguiente enlace:

[Bibliometrics Scielo Looker Studio](#)

Las tablas y gráficos que ilustran y evidencian los resultados obtenidos a través del análisis bibliométrico fueron generados por medio de un *dashboard* interactivo, co creado con un profesional especializado en analítica de datos, a quien conocí durante mi trabajo en GobLab<sup>29</sup>. Cualquier usuario con acceso al visualizador puede analizar los datos a su manera, de forma muy amigable, a través de apenas algunos clics. El *dashboard* genera frecuencias, porcentajes, análisis de tasa de crecimiento y relevancia de términos, entre otras funcionalidades pre concebidas y diseñadas en el proceso de incorporación de los datos. La gracia de este recurso es el hecho de que la mayor parte de los grafos son clicables y se adaptan a la correlación entre las variables que la persona desea conocer. Esto permite mayor variabilidad de análisis de los datos, combinación de análisis y múltiples interpretaciones. Así, recomiendo fuertemente que el lector tome unos minutos para conocer e interactuar con

<sup>29</sup> En 2022 conocí a Izaack Amorim. por mi trabajo en el Gob Lab, y entendí que podría hacer algo similar con la base de datos que venía trabajando en Scielo para mi investigación. Lo contraté para confeccionar este dashboard. En tres días Izaack me devolvió el visualizador para el primer feedback. Dashboard elaborado a través de Looker Studio Google.

los hallazgos del estudio bibliométrico emprendido. Invitamos al lector a interactuar y descubrir nuevas posibilidades de análisis<sup>30</sup>.

El proceso de su creación, sin embargo, no es completamente automático. Por detrás de la aplicación, hay una tabla (base de datos Scielo) confeccionada y adaptada a las potencialidades pre existentes en Looker Studio. Como mencioné, en las secciones anteriores, la base de datos de publicaciones en big data en las ciencias sociales latinoamericanas en Scielo, fue hecha en una tabla Excel, de manera manual, a través de la categorización y clasificación de algunas de las informaciones contenidas en el texto. Esta tabla fue organizada para facilitar la implementación del dashboard a partir de ella.

Los resultados como frecuencia y porcentaje generados por el *dashboard* pueden de igual manera ser extraídos de automatizaciones de *Excel*, por ejemplo. O hechos a partir de la estructuración en programas como SPSS. Sin embargo la posibilidad de obtener un mayor grado de especificación de las variables (análisis de nivel micro) en correspondencia a una segunda o tercera categoría/variable, no es tan simple de hacer en los ejemplos anteriores, como a partir del visualizador que construimos. Como podrán apreciar, el dashboard posibilita un *zoom in* y un *zoom out* en los análisis de la información disponibilizada. Es posible obtener los datos sobre un texto específico, e incluso acceder para descargarlo, a través de un único clic, como también, acceder a información más genérica, como la extensión bibliométrica de la base de datos, tasa de crecimiento de las publicaciones, tendencias disciplinarias, entre muchas otras.

La elección de presentar los datos con el visualizador de Looker Studio se da por la posibilidad que esta herramienta tiene de proporcionar visualizaciones personalizadas, creando gráficos, tablas y visualizaciones interactivas para representar los datos de una manera clara y comprensible, sobretodo para un usuario externo al ambiente de la investigación. una segunda justificativa es el hecho de que permite a los usuarios e interesados acceder a los datos bibliométricos en tiempo real, de manera personalizada (que muestran los indicadores bibliométricos más relevantes para el proyecto), lo que permite una visualización concentrada y rápida de la información más importante. Una tercera razón, es que facilita la colaboración: permite compartir paneles y visualizaciones con colegas, colaboradores o audiencias interesadas, facilitando la comunicación y el análisis conjunto, e

---

<sup>30</sup> link: <https://lookerstudio.google.com/reporting/ebea3775-01c9-4e0a-9b7d-7b51f5a59082>

incluso posibilita que los conjuntos de datos pueda ser escalado para adaptarse a las necesidades cambiantes del proyecto, o a la incorporación de nueva información.

Además de que Looker Studio permite que los usuarios exploren los datos por sí mismos, filtrar información y hacer preguntas específicas sobre los datos bibliométricos. Así justifico esta decisión por otras razones, como por ejemplo la posibilidad de proporcionar un mejor entendimiento y descubrimiento de patrones, tendencias y especificidades presentes en este estudio.

## **3.2 Resultados generales**

### ***3.2.1 Extensión bibliométrica***

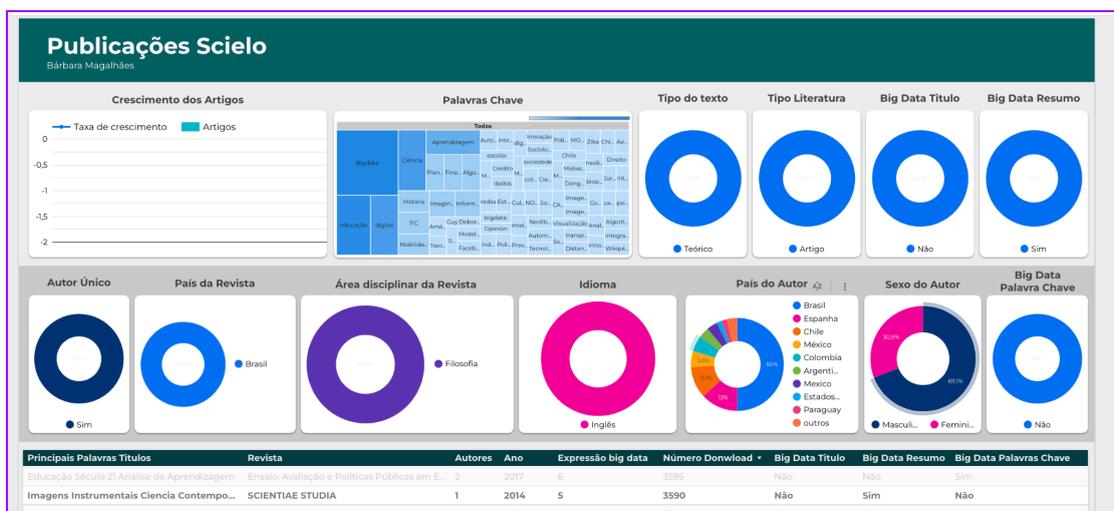
El total de publicaciones en big data en textos de “ciencias humanas” de revistas latinoamericanas en Scielo fueron de 26, hasta diciembre de 2020. El año que se hace la primera publicación es en el año de 2014. “On the status and role of instrumental images in contemporary science: some epistemological issues<sup>31</sup>. Un texto en inglés escrito por un autor portugués en una revista brasileña de Filosofía: “Scientiae Studia” volumen 12. Contaba con 3590 descargas. Un artículo de tipo teórico, conteniendo 5 repeticiones del término “big data” uno de ellos en el resumen.

Abajo (figura 14) se expresan los principales atributos proveniente del análisis individual del texto que marca la entrada del término big data en las publicaciones latinoamericanas en las ciencias humanas:

---

<sup>31</sup> <https://doi.org/10.1590/S1678-31662014000400002>

**Figura 14.** Atributos primer texto en CS y big data, 2014



El ejercicio de interpretación y análisis demostrado en el párrafo anterior, fue repetido de igual manera a cada texto, uno por uno. El objetivo de esta actividad era ir familiarizando aún más con la base de datos, y a través de la observación generar otras ideas, correlaciones y posibles resultados, en un intento de probar y jugar con los datos en general, pero también ir dando a conocer las especificidades de cada texto.

Esta es justamente una de las ventajas del *dashboard*: más allá de ofrecer una panorámica visual de los datos de fácil interpretación, es también permitir una interacción amigable y simple a los datos, de manera individualizada (observe que en cuadro verde de más abajo, uno tiene la posibilidad de seleccionar individualmente los textos y incluso acceder individualmente los archivos digitales de cada texto, para la lectura individualizada) lo que me permitió analizar las métricas de texto por texto más atentamente. Proceso que facilitó y mucho mi interacción con los autores de la publicaciones en ocasión de las entrevistas futuras.

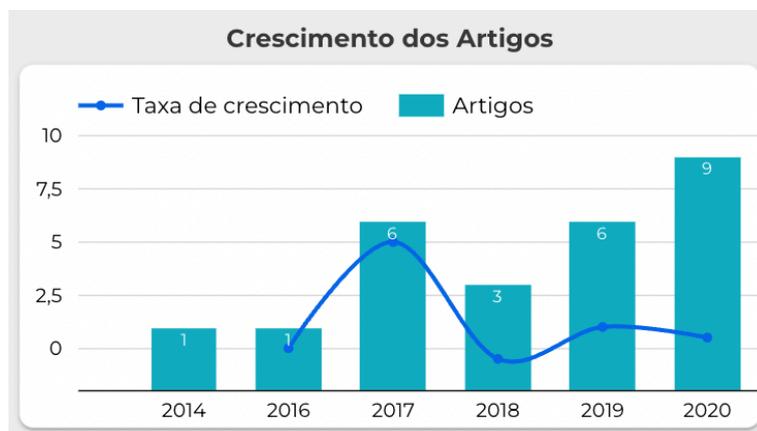
En la secuencia, sin embargo, presentaré los resultados de manera agrupada, o sea, pertenecientes al conjunto de los textos estudiados. Caso el lector tenga interés de conocer las especificidades de cada publicación recomendamos acceder al *dashboard* para tanto.

### 3.2.2 Tasa de crecimiento

Como se puede notar en la figura 15 el año con mayor número de publicaciones fue el de 2020, con un total de 9 textos. Sin embargo el periodo con la mayor tasa de crecimiento de

publicaciones fue en 2017. 96% de la literatura está constituida de “artículos”, siendo que 18 de ellos (69,2%) son de tipo teórico y 8 (30,8%) son de tipo metodológico.

**Figura 15.** *Publicaciones en BD por cada año + tasa de crecimiento*



### 3.2.3 Textos metodológicos

Estas dos últimas categorías (tipo metodológico y tipo teórico) consisten en una interpretación personal de los contenidos de los textos, con base a lectura individualizada de cada uno de las publicaciones. La identificación y distribución de estas categorías no solamente me obligó a releer los textos de forma individual, como también, al final de este proceso, me proporcionó el entendimiento (y hallazgo) de que en las ciencias sociales/ciencias humanas estamos aún discutiendo las implicaciones, los por venir y los posibles impactos de la utilización o publicación de tecnologías big data y del procesamiento de grandes cantidad de datos, llevando a un segun entendimiento: de hecho somos muy pocos aplicando literalmente tecnología big data en sus investigación.

Los textos de tipo metodológico, o sea, más “manos a obra” o que presentan resultados de investigaciones ya realizadas o en curso, son provenientes en su mayor parte de revistas de Colombia (4 textos - 50%), seguidos de Brasil (3 textos - 37,5%) y México (1 texto - 12,5%). Un cuarto de ellos presenta el término “big data” en el título y la mitad también en el resumen.

Ellos contabilizan un total de 105 utilizaciones de la expresión “big data”, un 40% comparado al total de 257 términos encontrados en estas 26 publicaciones, como está evidenciado en la tabla 3.

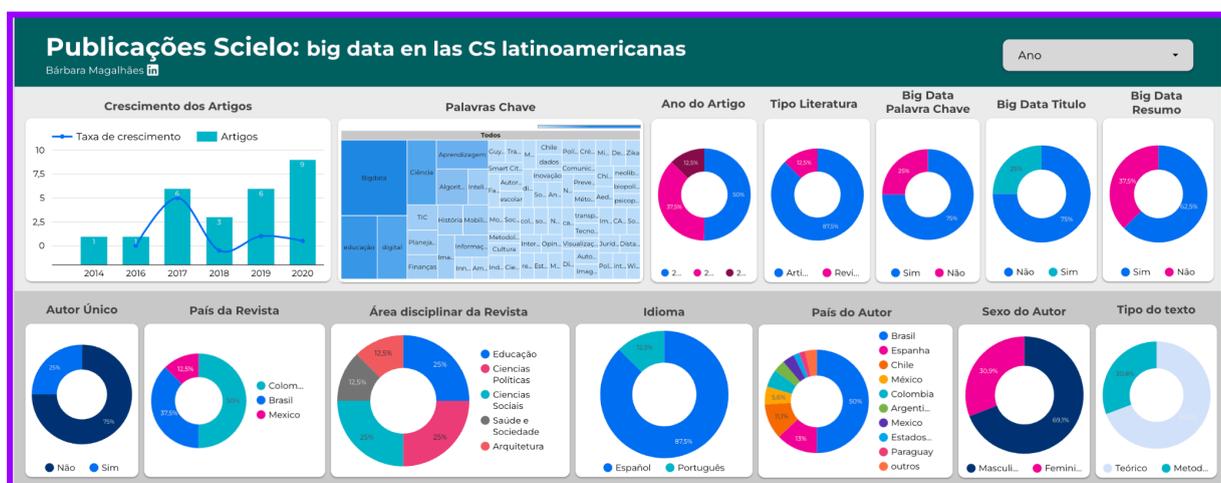
**Tabla 3.** Total de términos BD por tipo de texto

Total de términos BD en las 26 publicaciones	Total de términos BD Textos Metodológicos	Total de términos BD Textos Teóricos
257	105	152
cantidad de textos	8 textos (30,8%)	18 textos (69,2%)

Como se puede ver, aunque el total de términos big data sea más proeminente en textos de tipo teórico, en proporción, observemos que los textos metodológicos son en realidad los que más la utilizaron, pese a la menor cantidad de textos en esta categoría. La concentración de la palabra big data en los textos metodológicos nos evidencia una mayor y real incorporación de trabajos que utilizan o trabajan con este término, en publicaciones que efectivamente están usando este tipo de datos masivos, interés de esta investigación.

Otro dato: 75% de ellos son escritos por más de un autor, lo que demuestra una mayor necesidad de trabajar en conjunto y colaboración en publicaciones más ejecutivas, comparadas a los textos de tipo teórico.

Las publicaciones de tipo metodológica, están también escritas en su mayoría en español (87,5%). Las áreas disciplinares de las revistas que más se destacan en los textos metodológicos son “educación”, “ciencias políticas” y “ciencia sociales”, siendo que 50% de ellos, fueron publicados en el año de 2020, Como es posible observar en la figura abajo:

**Figura 16.** Datos generales: textos de tipo metodológico

### 3.2.4 Textos Teóricos

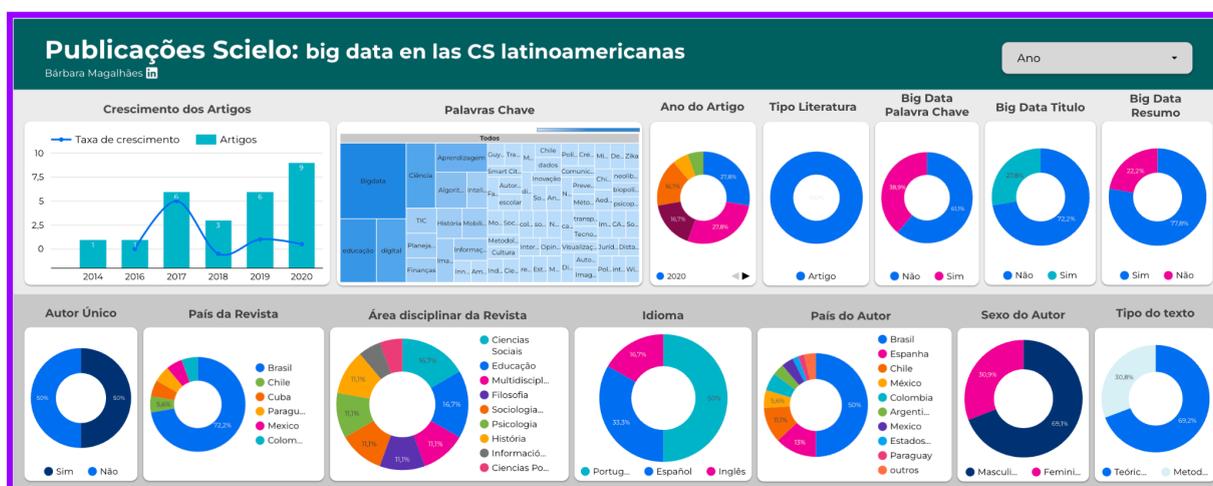
Los artículos clasificados como teóricos, en su mayoría fueron escritos en el idioma portugués (50% - 9 textos), seguidos del español (33,3% - 6 textos) y el inglés, con 3 textos, representando un 16,7% del total.

Así como sucede con los textos metodológicos, las áreas disciplinares “ciencias sociales” y “educación”, son más también las más frecuentes en los textos teóricos. Con tres publicaciones cada una, lo que representa apenas el 16,7% del total, seguidos de todas las demás áreas. Al contrario de los textos metodológicos, las publicaciones con contenido más teórico, tienden a poseer una distribución por área de disciplina bien pareja: dos textos por área disciplinar.

La revista que más publica entre los países de latinoamérica es de Brasil (72,2%, 13 textos de los 18 en total). Todos los otros países poseen 1 publicación cada uno. Esta tendencia se repite con respecto a los países del autor: Brasil en primer lugar con 50% de ellos, seguidos por España (13%), Chile (11,1%) y México (5,6%)

Del total de los 18 textos teóricos, la mitad de ellos fueron escritos por un autor único, la otra mitad en colaboración con por lo menos más de un autor. 77,8% de las publicaciones incluyen el término “big data” en sus resúmenes, porcentaje muy parecido a los que poseen el mismo término en los títulos (72,2%) y palabras claves (61,1%). La siguiente figura ilustra y respalda los resultados presentados:

**Figura 17. Datos generales: textos de tipo metodológico**



A partir de la categorización de los textos tipo teórico y tipo metodológico, fue posible constatar que las publicaciones más abundantes, las teóricas, se centran en la construcción y desarrollo de conceptos, marcos teóricos o modelos más abstractos. A diferencia de los textos empíricos, que se basan en la recopilación y análisis de datos concretos, los textos teóricos se dedican a explorar y explicar relaciones, patrones y fenómenos desde una perspectiva conceptual, analítica, y muchas veces, al menos en esta base de datos, de inferencia.

Lejos de ser una debilidad, la mayor presencia de publicaciones de este tipo demuestra una realidad en la actualidad con respecto al universo de las investigaciones en ciencias sociales en big data: los textos teóricos han sido esenciales para explorar nuevos conceptos y desarrollar marcos teóricos sólidos en un campo de estudio en evolución. Esta constatación va de la mano con los resultados obtenidos a través de la generación del mapa nube de palabras (figura 12) anteriormente presentado: los términos en mayor evidencia, como “datos”; “inteligencia”, “estudiantil”, “protección”, “opinion” y “derecho” evidencian el carácter más exploratorio de los textos. Estos son términos más generales, investigativos y menos técnicos: lo que nos esclarece sobre la temática mayoritariamente presente en las publicaciones, que aún no exploran conceptos técnicos y prácticos relacionados a big data.

Esto no lleva a otra conclusión: la evidente escasez de datos empíricos en los textos investigados. En el ámbito de los datos masivos, puede ser desafiante obtener y analizar datos empíricos de alta calidad debido a la magnitud y complejidad de los datos disponibles. Como resultado, los investigadores pueden optar por centrarse en la elaboración de teorías que ayuden a comprender los posibles patrones y relaciones dentro de esos datos, en lugar de realizar análisis empíricos exhaustivos. A parte de esto está el hecho de que la manipulación de estos tipos de datos requiere una especialización y capacitación específica que, al parecer, todavía no ha llegado a los profesionales y académicos en ciencias sociales. Indagaré sobre ello en los capítulos etnográficos.

Aun así, considero que esta realidad es pertinente al proceso de madurez de un campo disciplinario todavía en desarrollo, es decir, emergente. En este escenario investigativo es crucial establecer una base sólida de teorías y enfoques antes de adentrarse en la recolección de datos. Los textos teóricos permiten a los investigadores definir términos clave, identificar lagunas en el conocimiento, tantear y experimentar de manera más exploratoria en un campo en consolidación y establecer una dirección clara para futuras investigaciones empíricas. Esto es lo que veo que está pasando.

Por fin, otras tres dimensiones posibles de capturar a partir de la lectura y análisis de las publicaciones de tipo teórico es a) la interconexión interdisciplinaria, o sea, el análisis de big data en ciencias sociales a menudo requiere una combinación de conocimientos de diferentes disciplinas, de modo que los textos teóricos pueden servir como un puente entre disciplinas, al proporcionar un lenguaje y un marco conceptual en común para abordar cuestiones multidisciplinarias; también b) la generación de hipótesis y orientación de investigación: los textos teóricos acaban por generar nuevas preguntas de investigación y formular hipótesis que posteriormente pueden ser sometidas a pruebas empíricas. Estas teorías también pueden proporcionar orientación sobre qué tipos de datos buscar y cómo interpretarlos; y c) el enfoque en tendencias y prospectiva: dado que el campo de los macrodatos está en constante evolución, los textos teóricos pueden abordar tendencias emergentes, desafíos metodológicos y cuestiones éticas, guiando así futuras investigaciones e implementaciones efectivamente prácticas.

### 3.2.5 Autores

Con respecto a la característica de los autores, sabemos que dentro los 26 textos, se encontró un total de 55 autores. Una media de 2,5 autores por publicación. Sobre la colaboración en autoría, el número de publicaciones que poseen autor único (42,3% - 11 textos) es ligeramente menor que los que poseen 2 o más autores (57,7% - 15 textos). La presencia de múltiples autores en las publicaciones indica una tendencia hacia la colaboración en el campo de las ciencias sociales. La mayoría de las publicaciones involucran a dos o más autores, sugiriendo la importancia de la cooperación en la generación de conocimiento en el ámbito del análisis de big data.

Brasil es el país con mayor número de autores que publican sobre big data en las ciencias sociales en latinoamérica. Corresponde a la mitad de todos los textos (50%), seguidos de España (13%), Chile (11,1%), México y Colombia (5,6% cada uno).

**Tabla 4.** Cantidad de autores por país de publicación

País	Cantidad de autores
Brasil	27 (50%)
España	7 (13%)
Chile	6 (11,1%)
México	3 (5,6%)

Colombia	3 (5,6%)
Argentina	2 (3,7%)

El dominio de Brasil en este tipo de investigación refleja el enfoque significativo del país en este campo. La prominencia de autores brasileños como los principales contribuyentes en la investigación de big data en ciencias sociales en América Latina señala un liderazgo en la producción científica relacionada con big data en la región.

2014 y 2016 son los únicos años cuya colaboración con otro autor no fue predominante en las publicaciones estudiadas. Con un texto por cada año, sus publicaciones posee autor único. A diferencia de todos los demás años (2017, 2017, 2019 y 2020). 2020 es el año con más número de publicaciones hechas en coautoría: 77,8% del total de textos, siendo que la mayoría de estos fueron escritos por 3 autores. La evolución de la colaboración a lo largo de los años, con un incremento en 2020, sugiere un aumento en la importancia de la interdisciplinariedad y la colaboración en la generación de conocimiento. Este aumento puede estar vinculado al reconocimiento de la necesidad de abordar cuestiones complejas relacionadas con el big data.

Los textos de autores únicos son en su mayoría (81,8%) de tipo teórico y poseen la distribución de países por revista bien pareo, con una leve superposición de Brasil ante Colombia, seguido de Cuba, Paraguay, México y Chile. Los que poseen escritura colaborativa, no obstante son menores, 60% de textos teóricos, distribuidos en apenas 3 países de revistas; Brasil, Colombia y México.

Publicaciones con autores únicos son más frecuentes en el idioma español (63,3%), mientras que los textos con más de 1 autor está escrito en portugués, 46,7%, seguidos del español e inglés.

El 100% de la publicaciones de esta base de datos proviene de autores vinculados a instituciones académicas, más específicamente a universidades alrededor del mundo, lo que subraya la naturaleza científica de las investigaciones y la importancia de la academia en la generación de conocimiento relacionado con el big data en ciencias sociales. Los autores españoles (7) son los únicos provenientes de instituciones de fuera de latinoamérica. Aun si, la mayoría de los autores provienen de América Latina, con una notoria contribución de autores españoles que pertenecen a instituciones fuera de la región.

### 3.3.6 Género

De los 55 autores 38 de ellos son hombres (69,1%), 17 mujeres (30,9%). Aunque la diferencia entre hombres y mujeres sea importante, pude percibir que en 79% de los textos colaborativos, o sea, con dos o más autores, son mujeres las que firman como primera autora. Aunque la representación de mujeres como autoras es menor en general, la colaboración entre mujeres en textos con dos o más autores es notablemente alta, sugiriendo una dinámica diferente en términos de liderazgo y participación en proyectos colaborativos. Esto podría reflejar una tendencia hacia roles más activos con respecto a liderar el proceso de difusión científica o la escrita de los documentos. La observación de que las mujeres son más propensas a asumir el papel de primera autora en colaboraciones puede señalar que las mujeres están liderando el proceso de difusión científica o que están asumiendo roles más ejecutivos en proyectos colaborativos. Esto plantea cuestiones relacionadas con las dinámicas de poder y género en la colaboración científica.

De los 11 textos de único autor, apenas 3 son mujeres, mientras que 8 de ellos son de hombres. La mayoría de las colaboraciones de tres o más autores son entre hombres, y no hay ninguna publicación que haya sido escrita entre dos o más mujeres, mientras que entre dos o tres hombres ocurre 4 veces. La distribución de Género en Autoría Única puede evidenciar una desigualdad de género o un indicio de barreras adicionales que las mujeres enfrentan en la participación individual en la investigación.

Los textos con predominancia de mujeres como autora (3 textos - 11,5% del total), apenas uno de ellos es de tipo metodológico. lo que demuestra una tendencia de los textos escritos por mujeres serían de naturaleza teórica, mientras que 8 publicaciones son escritas por apenas hombres o mayoría de hombres.

La tabla a seguir evidencia, más detalladamente la diferenciación de publicaciones firmadas por mujeres y hombres:

**Tabla 5.** *Comparación publicaciones entre hombres y mujeres*

total autores	n° autores mujer	n° autores hombres	+ de 1 autor & 1° autor mujer	+de 1 autor & 1° autor hombre	autora única mujer	autor único hombre
55	17	38	6	9	3	8
mujer mujer	hombre hombre	textos predominan	textos predominan			

		<b>cia de mujeres autoras</b>	<b>cia de hombres autoras</b>			
0	4	3	8			

### 3.2.7 Colaboración:

Con respecto a la distribución de autores por publicación, percibirse que la distribución de autores es equitativa, con un 57,7% de los textos escritos por más de un autor. Esto sugiere una ligera tendencia hacia la colaboración y el trabajo en equipo en la generación de conocimiento en el campo del análisis de big data en ciencias sociales.

Referente a la frecuencia de las colaboraciones, es posible notar que de los textos escritos por más de un autor, 8 de ellos son el resultado de la colaboración entre tres autores, seguido por 4 textos con dos autores y 3 textos con cuatro autores. Esto demuestra una distribución variada en términos de la cantidad de colaboradores involucrados en las investigaciones. La tabla abajo señala la especificidad de la colaboración entre autores en la base de datos estudiada.

**Tabla 6.** Colaboración entre autores

Cantidad de Autores	Nº publicaciones	Porcentaje del total
Total	26	100%
Más de 1 autor	15	57,7%
Autor Unico	11	42,3%
2 autores	4	15,38%
3 autores	8	30,7%
4 autores	3	11,5%

Aunque la colaboración entre autores sea más común, un 42,3% de las publicaciones son el trabajo de un autor único, mostrando, de igual manera, la importancia de la autoría individual en el proceso de generación de conocimiento en el ámbito de las ciencias sociales y el análisis de big data. Aunque los datos indican una preferencia por la colaboración en las publicaciones, también muestran que una proporción considerable de textos son producidos

por autores únicos. Esto revela una rica variedad de enfoques en la generación de conocimiento, con un equilibrio entre la colaboración y la autoría individual en el campo del análisis de big data en las ciencias sociales.

### 3.2.8 Lista de textos más relevantes (más descargados)

En consideración a los textos en ciencias humanas y big data más relevantes para la audiencia, les presento la lista de las 8 publicaciones más descargadas en la plataforma:

**Figura 18.** Listado de textos más descargados en la plataforma.

Principais Palavras Títulos	Revista	Autores	Ano	Expressão big data	Número Download	Big Data Título	Big Data Resumo	Big Data Palavras Chave
Sociologia Digital Século 21	Sociologias	1	2016	23	24696	Não	Não	Sim
Bigdata Algoritmo Ciências Sociais Redes dig...	Cinta de Moebio	1	2019	19	9525	Sim	Sim	Sim
Bigdata Mídias Sociais Redes Gestão	Saude e Sociedade	3	2017	8	6674	Sim	Não	Sim
Neoliberalismo Psicopolítica Transparencia	Psicologia & Sociedade	1	2017	5	6089	Não	Sim	Não
Bigdata Propaganda Industria cultural	Psicologia USP	2	2018	26	5844	Sim	Sim	Sim
Desenvolvimento Tecnológico Policial Preven...	Revista de Relaciones Internacionales, Estr...	1	2020	5	5137	Não	Sim	Não
Educação Século 21 Análise de Aprendizagem	Ensaio: Avaliação e Políticas Públicas em E...	2	2017	6	3595	Não	Não	Sim
Imagens Instrumentais Ciencia Contempor...	SCIENTIAE STUDIA	1	2014	5	3590	Não	Sim	Não

*Nota.* Para mejor visualización de los términos, hacer zoom en el documento. Visualización recomendada 150%.

De los 5 textos más descargados, 4 de ellos están escritos en portugués, y el de la posición 5, en español. De los 8 autores que firman estas publicaciones, 7 son brasileños y uno proviene de Chile. Como se puede ver en la figura 18, dos de ellos son de revistas en “ciencias sociales”, dos en “psicología” y una en “salud colectiva”.

Abajo se ilustran las principales palabras/términos presente en los títulos de las 8 publicaciones más relevantes:

**Figura 19.** Principales términos asociados a las publicaciones más relevantes.



Como podemos percibir, big data es obviamente el término con mayor frecuencia en los títulos de tales publicaciones y está más fuertemente asociada a la digitalidad (perciba como el término digital aparece en el imagen de maner sobresaliente y muy cercana a “bigdata”), pero también a “algoritmos” y “opinión”.

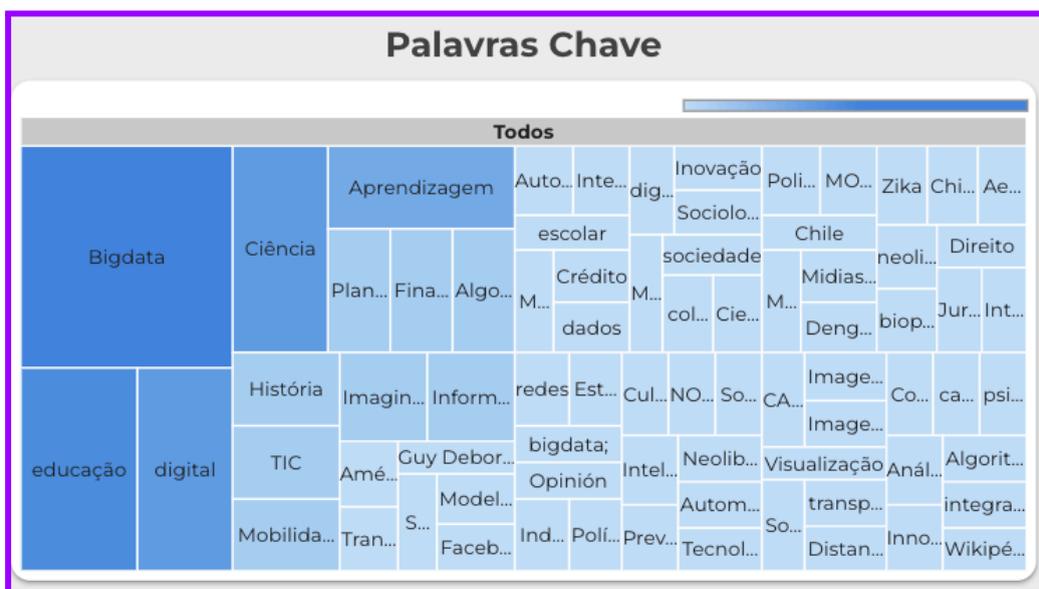
Otro término que se repite con preponderante frecuencia es “redes sociales” y “cultura” también relacionada a “psicopolítica”, “política”, “neoliberalismo” y “biopolítica”, evidenciando una fuerte preocupación con temáticas sociales relacionadas al interés de los estudios sociales y de del control social. Hay una tendencia en vincular la tecnología big data a cuestiones de digitalización de la vida, control de población por medio de la utilización de las redes sociales, en un mundo neoliberal, industrial y tecnológico. De igual manera podemos percibir también cuán cerca está el término “sociología” del “digitales”, evidenciando una asociación entre los estudios de la digitalidad, aparatos digitales, digitalidad al interés de la sociología alrededor de big data.

Aún con respecto a las palabras claves más utilizadas en las publicaciones analizadas, la figura a seguir evidencia los términos más frecuentes presente en el campo “palabras claves” de todas los textos en ciencias humanas en Scielo. Fue capturada a través del conjunto de

palabras provenientes de todas las publicaciones de la base de datos construida en este estudio bibliométrico.

Sin ninguna sorpresa, el término “big data” aparece como el más frecuente. contabilizando un total de 13 repeticiones. El segundo colocado es la palabra “educación” con 13 repeticiones, seguidas de “Ciencia” e “digital”, ambas con 5 palabras. aprendizaje: 4; Algoritmo: 3.

**Figura 20.** *Proporción de palabras claves en Scielo*



**Tabla 7.** *Frecuencia de términos más comunes de la base de datos*

Frecuencia	Términos
13	big data
6	Educación
5	Digital y Ciencia
4	Aprendizaje
3	Algoritmo
2	Inteligência Artificial TIC História Planejamento Movilidad Finanzas Imaginación Información

### 3.2.9 Idiomas:

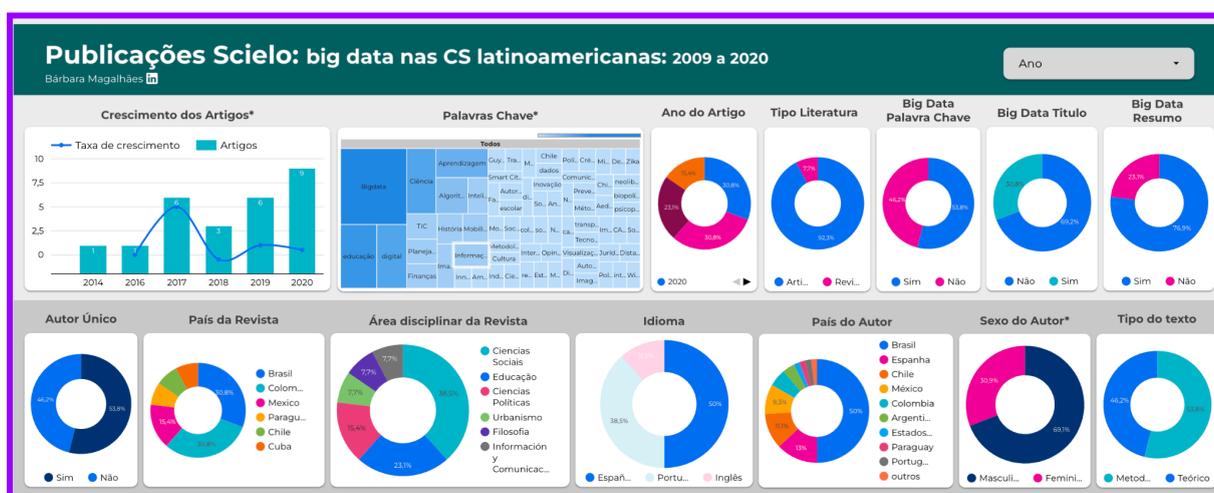
Como ya mencionamos, de los 26 textos totales, 50% de ellos están escritos en español (13 publicaciones), seguidos del Portugués, representando 38,5% con sus 10 publicaciones y Inglés con apenas 3 ejemplares, un 11,5% sobre el total de material analizado. Así, el idioma principal de los textos es el español, con una equilibrada distribución entre enfoques teóricos (6 textos) y metodológicos (7). Del total de textos en español, un 69,2% no incluyen "big data" en el título, mientras que el 76,9% de ellos hacen referencia a este concepto en el resumen.

En cuanto a la colaboración, se observa un equilibrio en el tipo de autoría. Un 53,8% de los textos son producto de autores únicos, mientras que el 46,2% restante involucra la colaboración de más de un autor (6).

En términos de procedencia geográfica, los países con mayor representación en las revistas de idioma español son, en orden, Brasil, Colombia y México. En el análisis de las disciplinas abordadas, las ciencias sociales (5 - 38,5%) superan en representación a la disciplina de la educación (3 - 23,1%).

En relación a la nacionalidad de los autores, se identificó que Brasil (27 - 50%), España (7 - 13%) y Chile (6 - 11,1%) son los países con mayor presencia en términos de representación por autor.

**Figura 21.** Índices de las publicaciones en español



En relación a las publicaciones en idioma portugués, se destaca que de un total de 10 artículos, el 90% de ellos fueron escritos en revistas brasileñas, con una excepción en la revista colombiana *Literatura: teoría, historia, crítica*.

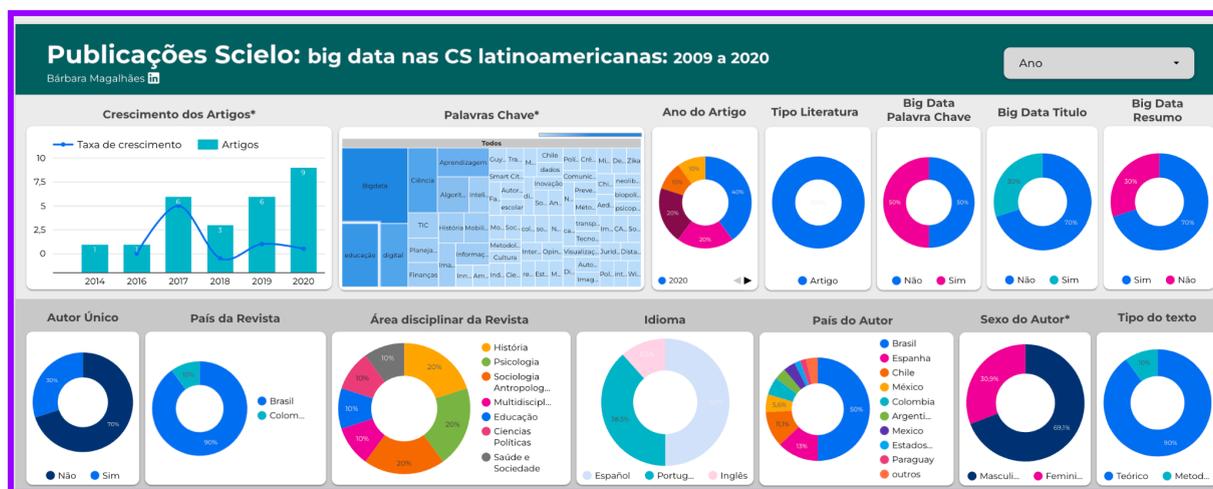
En términos de disciplinas, historia, psicología y sociología son las áreas que presentan un mayor número de publicaciones, con 2 artículos cada una. Además, un 90% de los textos escritos en portugués son de carácter teórico.

En cuanto a la procedencia de las publicaciones, se evidencia que Brasil y Colombia son los únicos dos países de origen de las revistas que contienen textos escritos en portugués. Sin embargo, existe una marcada disparidad entre ellos, ya que el 90% de los textos provienen de revistas brasileñas, mientras que solamente uno de ellos se origina en una revista colombiana.

En relación a la autoría, se observa que un 30% de las publicaciones en portugués son atribuibles a un único autor, mientras que el 70% restante fue elaborado por dos o más colaboradores.

Entre los 55 autores de las publicaciones analizadas, se destaca que el 50% de ellos provienen de Brasil (27 autores), seguidos por autores españoles con un 13% (7 textos) y autores chilenos con un 11,1% del total. Estos tres países representan los principales contribuyentes en términos de publicaciones en idioma portugués. Respecto a los años de publicación, se identifica que el año con mayor número de publicaciones en portugués es el 2020. Se trata principalmente de artículos de literatura que incluyen, al menos en el título y el resumen, el término "big data" en un 70% de los casos.

**Figura 22.** Índices de las publicaciones en portugués



### 3.2.10 Áreas disciplinares de la revistas:

Con respecto a los resultados de las áreas disciplinares de las revistas (11 distintas disciplinas), “ciencias sociales” y “educación” son las áreas que cuentan con mayor representación. Ambas poseen 5 textos cada una, lo que equivale al 19,2% de las publicaciones en ambos casos. En la secuencia, “ciencias políticas” se posiciona con 3 textos, representando el 11,5%. Las restantes áreas del conocimiento presentan 2 ejemplares cada una, sumando un 7,7% del total de publicaciones.

En el ámbito de las ciencias sociales, todos los textos fueron escritos en español y principalmente provienen de revistas brasileñas (3), seguidos de una publicación de Paraguay y otra de Chile. En esta área, la colaboración es común, con un 60% de los textos escritos en colaboración. Esta misma proporción se refleja en el tipo de texto, ya que el 60% de ellos son de carácter teórico (3 textos).

En el caso de la “educación”, los textos siguen predominando en su mayoría en el enfoque teórico. En contraste, en “ciencias políticas”, a pesar de no ser la disciplina más frecuente, es la que presenta la mayor cantidad de textos de tipo metodológico.

En ambas “ciencias políticas” y “educación”, las publicaciones en español son dominantes. Además, la historia se destaca como la que presenta la mayor cantidad de textos escritos en portugués, todos de carácter teórico.

Estos resultados proporcionan una comprensión más profunda de las preferencias temáticas y metodológicas en distintas áreas disciplinares de las revistas en ciencias sociales en la base de datos Scielo. La distribución de publicaciones por disciplina y su enfoque temático revela patrones interesantes sobre las tendencias de investigación en estas áreas.

## Cap 4: Scopus: ampliando el escopo de investigación

La decisión de trabajar con Scopus, fue, como otras variables de esta investigación, dada por el azar que Knorr-Cetina nos invita a atender (2005). Un día recibo el correo de correo de mi profesor guía: “Bárbara, acaba de salir la programación del ESOCITE-LALICS 2020/2021<sup>32</sup>. Creo que deberías ir. Hay una mesa sobre big data, te dejo el link” (comunicación personal, 2020). El pedido de un profesor guía es una orden. Era la primera vez que lo hacía como estudiante de doctorado e investigadora perteneciente al campo de los estudios de ciencia y tecnología.

En la segunda presentación a la que asisto ese día “Las tecnologías de big data y el sector público: un análisis bibliométrico”, aparece en pantalla el título en la primera lámina de la presentación del exponente Henry Chavez<sup>33</sup>. A esta altura, ya había comenzado el proceso de recolección de información de la base de datos Scielo.

**Figura 23.** Congreso ESOCITE- LALICS, mesa 31



<sup>32</sup> Referencia exacta: Primer Congreso ESOCITE - LALICS (XIII Jornadas). Los contenidos y registro de todas las actividades del Congreso pueden ser acompañadas a traves del canal de YouTube de lo mismo: <https://www.youtube.com/channel/UCSzIXZIR2M7zYIgMGLxU1Cw/featured>

**Esocite:** Asociación de estudios Sociales de la Ciencia y la Tecnología. **Lalics:** Latin American Network for Economics of Learning, Innovation, and Competence Building Systems.

<sup>33</sup> PhD in Social Sciences from the Ecole des Hautes Etudes en Sciences Sociales in Paris, France. Interdisciplinary background in social sciences, economics and politics, specialist in social studies of science, technology and innovation; anthropology of global systems; public policy design and data science. He has worked as a researcher and consultant in the private and public sectors, as well as in social organizations, NGOs and international organizations. He is currently a member of the CTS-Lab in FLACSO – Ecuador and Divergence in France” : La descripción del científico en la página web del laboratorio, accedido en marzo 2022.

La presentación tenía dos aspectos interesantes para esta investigación: el primero y obvio, el análisis bibliométrico, y el segundo, la dupla de términos: “big data” y “sector público”. Las lecturas de los textos de la base de datos Scielo categorizadas como metodológicos, tenían en común el hecho de que definían big data como un sinonimo de datos vinculados a los banco de datos de sectores públicos, de los más variados tipos de servicios gubernamentales o en alianzas entre agentes públicos y privados. Ya había pasado en mi radar, que por lo menos en latinoamérica, la utilización real de big data parecía estar aconteciendo dentro de las oficinas de los estados.

Así que durante el congreso, en abril, ya tenía más o menos claro sobre el panorama, e incluso ya tenía algunos resultados provenientes de la base de datos Scielo, que se fueron confirmando a lo largo de la presentación de colegas, pero en especial de Henry Chávez, pues él estaba haciendo lo que yo pretendía, con otra base de datos Scopus.<sup>34</sup> En su presentación demostraba algunos de los hallazgos que yo misma había producido a partir de Scielo, ilustrado con gráficos que puso a disposición en el *chat*.

Prontamente ingresé al enlace. Entré a la página y pude ver que los gráficos se movían, o se cambiaban de forma o resultados, conforme uno iba clicando en determinadas partes de la imagen. No es necesario decir que me quede aún más impresionada y con ganas de replicar la misma manera de representar mis datos.

No dudé en mandarle un mensaje por privado al final de su exhibición, otra ventaja más de los congresos en línea. Henry prontamente se ofreció para ayudar en lo que fuera preciso y hemos colaborado desde entonces. Después de contarme detalles de su investigación y de percibir que de hecho estamos trabajando básicamente bajo el mismo objetivo: comprender la circulación de conocimiento alrededor de big data en el mundo científico. La diferencia es que su enfoque era global. El mio, latinoamericano.

---

<sup>34</sup>Esta circunstancia me hizo acordar una de las anécdotas que contaba Knorr-Cetina al estudiar el laboratorio que investigaba. Al hablar de las “varios ordenes o niveles de selectividad” (2005, p.62), refuta la idea de que los procesos científicos puedan ser reproducidos de la misma manera en circunstancias diferentes, haciendo que sus procesos no se repitan. Para apoyar su argumento, en una nota de pie escribe: “Esto explica la ocurrencia de “descubrimientos” simultáneos por científicos que en realidad no se roban unos a los otros. Nótese que las instituciones científicas y las formas conocidas de control social en la ciencia pueden ser vistas como una estructura amplia para asegurar que las selecciones queden en gran medida rijas, y que las restantes se hagan de una manera similar, compatible y repetible” (Knorr-Cetina, 2005, p.63). Me vi en la misma situación: dos científicos diferentes, de lugares diferentes, hasta entonces completamente ignorantes sobre la existencia del otro, interesados y produciendo en el mismo contexto, investigaciones bastante semejantes.

Por otra parte, me presentó la metodología que había estado utilizando para analizar la librería de Scopus y explicó sobre la potencialidades de conducir un estudio en esta base de datos. En medio de la conversación decidimos que iríamos colaborar para que la versión y el formato de su investigación global pudiera ahora ser replicada en versión latinoamérica y con enfoque en ciencias sociales, ya que justamente uno de los principales objetivos de sus estudios postdoctorales era consolidar una red de investigadores sociales latinoamericanos en big data e inteligencia artificial.

La viabilidad de esta reproducción no solamente era posible, pero también sería un procedimiento más o menos simple, ya que la parte dura (“la estructura” por detrás del procesamiento de la información, los comandos y las programaciones que harían correr los datos para generar los resultados), estaban ya hechas en función de su propia pesquisa.

Fue también en esta ocasión que descubrí que la manera como estaba trabajando manualmente los datos, y lo que llevaba meses haciendo, podría, con “un poco” de conocimiento en programación, ser realizado en menos de una semana, a una escala mucho más grande y con mayor capacidad de producir más variedad de resultados. Comprensión que fue corroborada por mi en innumerables ocasiones, por medio de las entrevistas realizadas a los científicos sociales autores de las publicaciones de ambas bases de datos.<sup>35</sup>

Seguir con mi estudio se volvió inmensamente más interesante, porque podría contar con un *partner* de investigación, y conocía directamente a alguien que me podría enseñar a programar (soñaba yo), o por lo menos, a desarrollar capacidades suficientes para reproducir su investigación, al ejecutar la mía de manera autónoma. Además, ahora contaba con alguien que estaba movilizandoo contactos a favor de la “causa big data” en todo mundo, cosa que al final, facilitó enormemente la conducción de las entrevistas a los científicos que hemos emprendido durante el proceder de la etnografía.

Finalmente, se abrió la posibilidad de establecer redes de contacto y de seguir una metodología y protocolo de análisis bibliométricos de más una base de datos, lo que permitió mayor alcance en mi investigación, como también mayor rapidez en los procesos de

---

<sup>35</sup>Una de las autoras, al responder sobre lo que estaba haciendo en la actualidad, contestó que, a propósito de todo lo que habíamos estado conversando (tamaño y capacidad de procesamiento de la base de datos que analizaba), ella solía manipular tablas de *Excel* de más de 4 mil líneas, y que la última base de datos en la que estaba trabajando, de características similares, le tomó alrededor de 6 meses para uniformizar, preparar y finalmente analizar los resultados. La científica aseguraba que si supiera algo de programación o de las técnicas de procesamiento de datos, como su otro compañero (tercer autor del artículo), habría tardado semanas, y no medio año. También expresó su interés por hacer un curso básico de programación, y así trabajar los datos de manera más automatizada.

levantamiento y construcción de la información. Sin mencionar el hecho de que nuestras investigaciones, en conjunto, podrían colaborar de manera complementaria. Había, entonces, posibilidad de conducir análisis de enfoque global/macro y regional/micro al mismo tiempo, lo que presento como un importante hallazgo de la fase etnográfica del estudio, en el capítulo nueve.

#### **4.1 Scopus: la base de datos más grande en latinoamérica**

Al salir de nuestra primera reunión, me habían quedado bastante claro los motivos por los cuales incluir Scopus en mis estudios bibliométricos. De hecho, en la misma presentación de Henry encontré razones suficientes.

Primero, Latinoamérica es mucho mayor que Brasil. Mi experiencia como migrante a países hispano hablantes, ya me había permitido ver que no solamente en la cultura musical, o las teleseries y comidas, Brasil es una isla continental en pleno continente latinoamericano. Lo mismo sucedía con el mundo científico.

Pude ver durante mis primeros años de doctorado, cómo Brasil se queda fuera por nuestra historia colonial portuguesa, que nos implantó un idioma diferente de todos los demás países de la región, “El idioma doméstico”(García Canclini, 1980; 1992; 2003). En cierta medida, los brasileños estamos amontonados en nuestras propias bases de datos, protocolos y maneras específicas de escribir y difundir la ciencia (ver capítulo 4 - Scielo).

Por otro lado, Scopus estaba en boca de todos mis compañeros del doctorado, en las charlas que frecuentaba, y en los textos del curso. Brasil, a mi parecer, está más familiarizado e influido por las teorías y prácticas científicas europeas, particularmente francesas. Esto podría explicar cómo creado una librería científica y académica para sí, por lo que Scielo me habría bastado.

Si quería dar a conocer los procesos latinoamericanos, tenía que emprender una investigación en la base de datos más popular de la latinoamérica hispano hablante e incluir Scopus.

Elegir analizar la base de datos de Scopus se justifica, además de la razón ya explicada, por su calidad multidisciplinar, cubre un gran número de revistas y realiza una selección rigurosa de sus publicaciones (Chaparro-Martínez, et al.,2016).

En segundo lugar, es muy popular en el mundo científico, siendo una de las dos bases de datos científicas más utilizadas en la academia, lo que contribuye de manera decisiva para los estudios de análisis y evaluación de la ciencia (Codina et al., 2020). Tercero, posee un excelente apartado de análisis de información académica por medio de sus métricas, índices y funcionalidades para la evaluación y estudio bibliométrico al interior de su propia plataforma digital (Chaparro-Martínez et al., 2016). Finalmente, en detrimento a Web of Science, por ejemplo, Scopus es la base de datos que contiene más revistas del área de ciencias sociales y humanidades vinculadas, 9451 títulos contra 6500 de Web of Science<sup>36</sup> (Codina et al., 2020).

Para proceder a explorar la base de datos, junto a Henry Chávez establecimos un cronograma de actividades, acordamos reuniones esporádicas y a lo largo de estos encuentros hemos generado los resultados que se presentan en el presente trabajo.

#### **4.2 Parámetros e índices para la aplicación de una metodología bibliométrica compartida**

Como se ha anunciado, este trabajo es una reproducción a escala latinoamericana, y con foco en las publicaciones en ciencias sociales, del estudio bibliométrico realizado por Henry Chávez del *CTS- Lab - FACSO Ecuador*. Baso esta decisión no sólo por la ventaja útil que me proporciona, también porque creo que las investigaciones que nos anteceden influyen en la producción:

El carácter "artificial" de la herramienta más importante del científico, el laboratorio, reside en el hecho de que éste no es más que una acumulación local de materializaciones de selecciones anteriores. Las selecciones de investigaciones anteriores también afectan las selecciones subsiguientes, modificando las condiciones de las nuevas decisiones. En consecuencia, los productos de la ciencia no sólo están impregnados de decisiones sino que también son impregnadores de decisiones, en el sentido de que señalan hacia nuevos problemas y predisponen a sus soluciones. (Knorr-Cetina, 2005, p. 64)

---

<sup>36</sup>Otro motivo muy relevante para no utilizar la librería WOS en este estudio bibliométrico tiene que ver con la imposibilidad de conducir tal investigación dentro de la legalidad del proceso. El acceso a este banco de datos como hicimos con Scopus, no fue posible básicamente por dos razones: los costos para adquirir la autorización excedían el presupuesto de esta investigación, lo que me imposibilitó de hacer un estudio bibliométrico semejante a los anteriores. La burocracia para acceder los responsables para un acceso libre, se volvió excesivamente complicada y costosa, lo que acabo imposibilitando la implementación de los estudios en WOS.

Las ventajas de reproducir el método bibliográfico desarrollado en el CTS-Lab Ecuador son varias: la primera de ellas es utilizar una metodología ya validada y establecida. Los resultados provenientes del estudio bibliométrico y emprendido por el *CTS-Lab Ecuador* llevan por lo menos 2 años en circulación. El hecho de que hayan publicado estos resultados, hace que la investigación sea más relevante, y sea también una evidencia, aprobada por una comunidad científica competente, al pasar por el ojo de los pares, a la hora de la sumisión del artículo.

En segundo lugar, ya estaba implementada operacionalmente. Esto significa que pude implementar paso a paso su metodología de obtención de información. Inclusive ocupé la misma base de datos, capturada, cualificada y procesada por Henry Chavez. Sus técnicas pueden revisarse en profundidad en “Big Data Research on Scopus, 2009-2019” de Chavez et al (2022). Aun si, les presento un resumen a título de clarividencia del procedimiento de levantamiento y procesamiento de información:

Hemos emprendido un estudio bibliométrico en la base de datos Scopus, de publicaciones que contienen la palabra clave “big data” en los título, resumen, o palabra clave en las publicaciones de ciencias sociales entre los años de 2009 a 2021 con la finalidad de identificar las principales tendencias, temas, redes e instituciones que publican sobre big data en todo el mundo y a explicar las relaciones y diferencias entre las comunidades científicas que trabajan sobre este tema en países centrales y periféricos, pero en este caso, más específicamente en países latinoamericanos.

Estos documentos de texto fueron compilados y luego validados usando rutinas de verificación<sup>37</sup> desarrolladas en Python 3.7 para excluir duplicados<sup>38</sup> y registros con información insuficiente (sin autor, ni afiliación) necesaria para realizar los. Luego se realizaron una serie de operaciones de división y clasificación de los registros con la finalidad de obtener una base “limpia” y operativa que priorizarse los registros a la escala individual por autor y publicación.

---

<sup>37</sup>Rutina de verificación: proceso automatizado que permite uniformizar la data de entrada y poder trabajar con ella a partir de información idealmente preestablecida Es necesario para la generación de información y resultados.

<sup>38</sup>Los duplicados son textos presentes en la misma base de datos más de una vez. Como el proceso de captura de las bases de datos es la etapa más manual de este procedimiento, es muy probable que errores humanos, tal como digitación, descargas duplicadas o posibles distracciones acontezcan. Así, para la identificación de estos errores, se utiliza un comando o script de programación para identificar este tipo de errores.

Para tener idea de la grandeza de la base de datos trabajada en esta etapa de investigación, hemos realizado un análisis bibliométrico de más de 73.000 publicaciones de este tipo entre 2009 y 2019. Las publicaciones de la base de datos Scopus que contienen la palabra clave "big data" se han disparado de 30 (2009) a casi 16.000 (2019). Esta tendencia revela la importancia y el interés por la temática en distintas disciplinas y contextos. Trabajos anteriores<sup>39</sup>, por ejemplo, también han analizado la aparición y las características de la investigación científica sobre big data, pero necesitan actualizarse.

Otro factor que demuestra la importancia de conducir nuestro estudio con una mirada latinoamericana, es el hecho de que las investigaciones previas ponen en relieve el real protagonismo de los investigadores y las instituciones chinas en este campo, junto con la influencia de las contribuciones estadounidenses, siendo las más citadas (Chavez et al., 2002).

Este estudio tiene como objetivo, por lo tanto, renovar las observaciones y descubrimientos presentados en investigaciones bibliométricas previas, llevando a cabo un análisis más amplio que abarque la totalidad del cuerpo científico de la última década. Además, a diferencia de la mayoría de estas publicaciones, que se han enfocado en evaluar el desempeño bibliométrico general del campo, nuestra investigación se centra en el análisis de las publicaciones latinoamericanas (y en comparación a las de todo el globo), en los textos producidos por científicos sociales de estos países.

Nuestro conjunto de datos se ha construido mediante una consulta general de documentos que contienen la palabra clave "big data" en los campos "title, abstract or keywords" de la base de datos Scopus. Esta consulta proporciona un resultado de 75.300 documentos publicados entre 1970 y 2019. A continuación, los resultados se exportaron como varios archivos '.csv' que contenían 2.000 entradas cada uno con toda la información disponible sobre citas y detalles bibliográficos, resúmenes y palabras clave, financiación y otros detalles. Estos archivos se compilaron y posteriormente se validaron mediante un algoritmo de verificación desarrollado en python 3.7 para excluir los duplicados y las entradas sin siquiera la información mínima (autor y afiliación) que se necesitaban para el análisis posterior, como ya mencionado un poco más atrás.

Nuestra base de datos fue sometida a un análisis en relación a tres aspectos principales:

---

<sup>39</sup> Gupta D, Rani R. (2019); Halevi G, Moed H.(2012); Parlina A, Ramli K, Murfi H. (2020); Raban DR, Gordon A. (2020); Singh VK, Banshal SK, Singhal K, Uddin A. (2015)

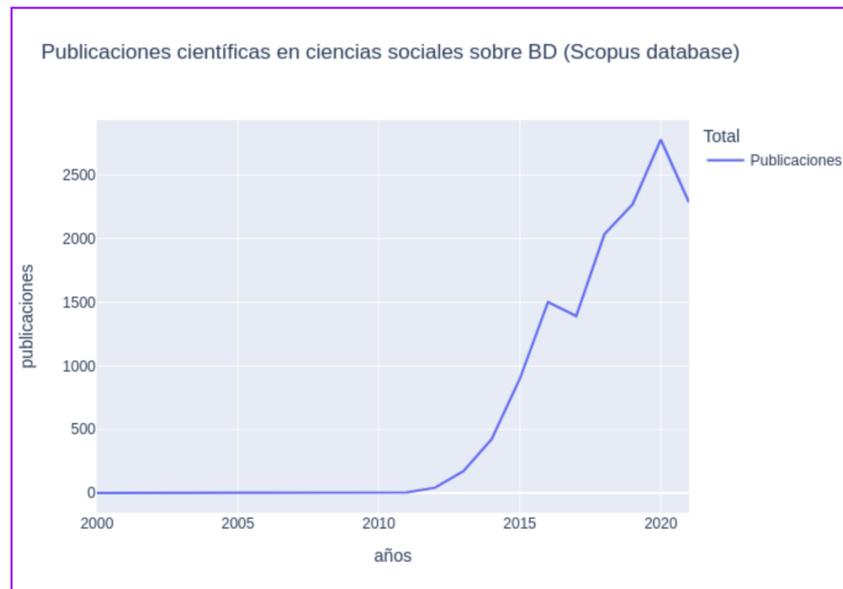
- *Indicadores de productividad y rendimiento científico:* Esto abarcó aspectos como el volumen de publicaciones, las tasas de crecimiento, los colaboradores más destacados, revistas relevantes y artículos categorizados por año, tipo de documento, instituciones y países.
- *Análisis temático:* Se evaluó la distribución de las publicaciones según áreas de investigación, las principales áreas de investigación por año y país, el análisis de palabras clave y la identificación de tendencias a lo largo del tiempo.
- *Redes de colaboración:* Se llevó a cabo un análisis de redes en tres niveles: autores, instituciones y países, con el fin de identificar las principales agrupaciones y tendencias tanto a nivel local como internacional en términos de colaboración.

Este enfoque integral permitió un análisis exhaustivo de la producción científica en el campo de estudio, destacando tanto los aspectos cuantitativos como cualitativos de la investigación.

#### **4.3 Resultados encontrados**

Los dos primeros artículos de ciencias sociales registrados en la base de datos Scopus que hace referencia a la palabra clave big data se publicaron en el año 2000. Sin embargo, después de esta publicaciones no se registran nuevos artículos sino hasta los años 2007 y 2009. sólo a partir de este momento se observa un número creciente de publicaciones sobre la temática que va a sostenerse hasta el 2020, cuando parece invertirse la tendencia. La Figura 24 abajo demuestra exactamente esta realidad: un crecimiento exponencial de textos sobre big data publicados en Scopus a partir de los años 2010 hasta apenas una década más. 2020 es el año donde las publicaciones experimentan una vertiginosa caída, tendencia que se observa hasta los días de hoy. El capítulo 7 hace una apreciación en profundidad sobre este hallazgo, confirmado, además, durante el proceso etnográfico de investigación.

**Figura 24.** *Número total de publicaciones en ciencias sociales sobre big data por año, 1970-2019*



Hasta diciembre de 2021, se han publicado un total de 13.820 artículos de ciencias sociales que contienen esta palabra clave o “concepto pre-coordinado”(Daban, RD. y y Gordon A., 2020) ya sea en el título, el resumen o la palabra clave de cada texto.

Aunque menos prolífica que la tendencia global, la fase de expansión de la producción científica en ciencias sociales siguió el mismo patrón exponencial de otras áreas de conocimiento. Pasó de menos de 10 artículos en 2010 a casi 2.800 a principios de 2020. Esto muestra una tasa media de crecimiento de alrededor del 85% anual.

Tal literatura (textos publicados en la base de datos Scopus, sobre big data en Ciencias Sociales) se compone principalmente de ponencias de artículos en revistas científicas (48%) y en conferencias (38%). Otras contribuciones son menos representativas: capítulos de libros (6%), revisiones (4%), libros (1%) y otros 2% (editoriales, notas, artículos en prensa, cartas, encuestas breves, etc.). En contraste con la producción global y en otras áreas, como ingenierías y ciencias médicas, donde las publicaciones en conferencias son mayoritarias (55%) frente a los artículos en revistas (33%).

La disparidad entre la frecuencia de publicaciones en artículos entre las ciencias sociales y las demás áreas, refuerza la conclusión a la que llegamos en el proceso de las entrevistas con muchos de los autores de estos textos: la necesidad de trasladar la difusión del conocimiento

social producido a partir de big data a formatos más libres y de mayor capacidad de actualización. Los textos académicos tipo artículos, para muchos de los entrevistados, sobre todo de áreas no afines a las ciencias sociales, son considerados obsoletos. Según ellos, los artículos tradicionales restringen la capacidad de representación y visualización de los datos, ya que cada vez más han surgido herramientas de automoción que dinamizan la presentación de la información producida. Esto requiere un soporte de hospedaje digital, capaz de conectar a través de enlaces y otras soluciones, la información producida, que muchas revistas ya poseen, pero que todavía no han adaptado sus reglas de publicación a los recursos “más tecnológicos” de escritura y presentación de datos. Este es el mismo caso de esta trabajo, además de la recomendación de lectura a través de un soporte digital, para garantizar la interacción del lector con los datos e informaciones producidas.

Otra potencialidad de los textos generados a partir de las conferencias, mencionado por los entrevistados, es la mayor facilidad para incorporar modificaciones y mejoras hechas por parte de la audiencia, en ocasión de las presentaciones a sus textos., se busca obtener retroalimentación temprana de la comunidad académica. La flexibilidad de este tipo de publicación, al parecer, se ajusta más adecuadamente a los estudios alrededor de big data y procesamiento automático y digital de datos, por la oportunidad de presentar resultados preliminares de investigación y de carácter más experimental, cosa fundamental para el desarrollo de investigaciones enmarcadas en este contexto.

Mirando desde el punto de las ciencias sociales, y si comparamos los resultados entre Scielo y Scopus, el hecho de que los autores en ciencias humanas están publicando más en formato de artículo, corrobora la información brindada a través del análisis de los datos en Scielo, de que tendemos a emprender investigaciones más tradicionales, teórica y exploratoria respecto de los conceptos, epistemologías y formas de trabajar que la introducción de big data a las investigaciones sociales pueden provocar.

Al analizar más detenidamente los textos de las conferencias, es posible percibir el carácter más metodológico de estos textos. La difusión de “los modos de hacer” de las investigaciones, es decir, de los métodos y metodologías, así como la discusión de resultados y difusión de *scriptis* y nuevos modelos para la resolución de determinado problema, son la tendencia más fuerte en este tipo de texto. “Las conferencias también suelen ser organizadas alrededor de un tema” dijo una de las entrevistadas. Esto hace toda la diferencia desde el punto de vista de trabajos con fuerte característica multidisciplinaria.

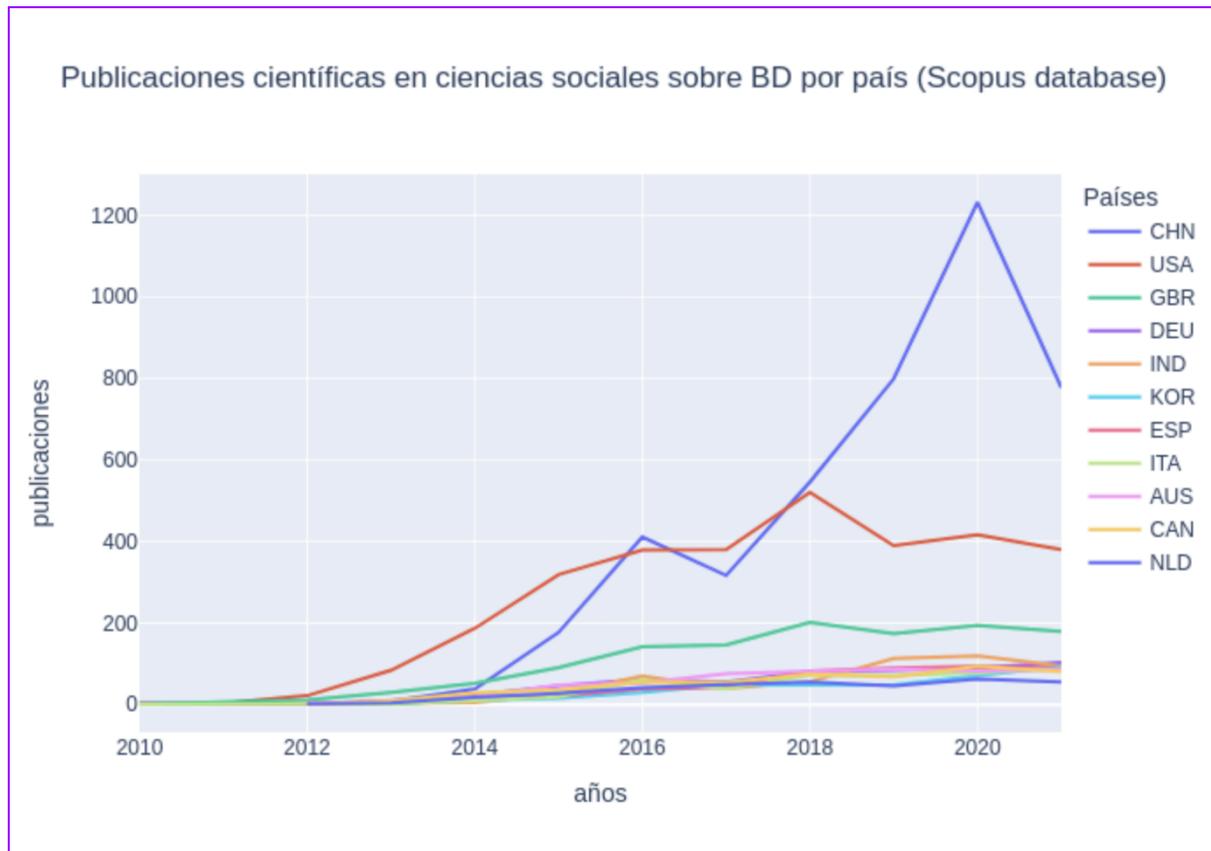
A menudo escuchamos que el desafío de publicar “cosas de big data” era la dificultad de encontrar revistas que reconocieran la temática trabajada pertenecientes al interés de la disciplina o campo de actuación al cual el periódico estaba asociado:

Este artículo que leíste, por ejemplo, sólo fue publicado porque era una conferencia de TI que hablaba en una mesa de trabajo específica, de metodología y nuevas tecnologías de aprendizaje. Antes habíamos tratado de publicar en dos revistas. La de arquitectura consideró que era demasiado de ciencias sociales para su enfoque, texto con foco en planificación y gestión del espacio público desde el punto de vista del tránsito; y la segunda, de sociología, consideró demasiado de urbanismo y tecnología para la línea de textos que la revista aceptaba [risas]. Es un poco angustiante y duro: hay que andar mirando mucho más posibilidades de revista, lo que torna más difícil y más trabajoso encontrar la apropiada. Siento que siempre que vamos a publicar sobre estos asuntos, a la hora de enviar el artículo, estamos dando un balazo en el oscuro (BRV - Entrevista 4/2022 - Brasil - Mi traducción).

De cualquier manera, la elección entre publicar en conferencias o revistas depende en última instancia de los objetivos del autor y la naturaleza de la investigación. Algunos investigadores pueden optar por ambas vías para maximizar la difusión de su trabajo, estrategia considerada ideal, pero casi imposible de ser ejecutada, por la gran mayoría de los entrevistados.

La producción científica alrededor de big data está distribuida asimétricamente alrededor del mundo. China (4.311) y Estados Unidos (3.085) mantienen un dominio indiscutible de este campo de estudios. Sin embargo, cabe resaltar que si bien los investigadores afiliados a Estados Unidos lideraron este campo durante sus primeros años de desarrollo, fueron rápidamente superados por los investigadores afiliados a China. Desde 2016, las publicaciones norteamericanas parecen haberse estancado por debajo de los 400 artículos por año, mientras que las chinas se han triplicado.

**Figura 25.** Publicaciones sobre big data por país con mayor número de publicaciones y año, 1970-2019



Una posible explicación del fenómeno es la alta probabilidad de que este rendimiento superior esté relacionado con la política tecno-nacionalista de I+D de China, cuyo objetivo es la transformación digital del país y de liderar mundialmente la industria de datos e inteligencia artificial (Fabre, 2018).

Gran Bretaña (1.224 textos), Alemania (560), Australia (546), India (528 textos) y otros países europeos también reflejan un interés por el tema. De hecho, tomados en conjunto, los países miembros de la Unión Europea suman 3.128 publicaciones, lo que los ubicaría por encima de los Estados Unidos.

Desde una perspectiva dinámica, al analizar la tasa promedio de crecimiento anual del número de publicaciones en los últimos cinco años, se observa que los Estados Unidos han mostrado un interés relativamente bajo en este tema, con un modesto aumento del 2%. En contraste, los países de la Unión Europea han demostrado un marcado interés, con un

crecimiento del 18%, mientras que Gran Bretaña, Australia y China también muestran cifras significativas de crecimiento del 12%, 10% y 28% respectivamente.

Sin embargo, la brecha se amplía aún más al considerar a las emergentes comunidades científicas de países como Rusia e Indonesia, que han experimentado un crecimiento excepcionalmente rápido en esta área desde 2015, con tasas de crecimiento del 52% y 51%, respectivamente.

Otros contribuyentes destacados, con un crecimiento anual de más del 40% desde 2015, incluyen a Sudáfrica, Vietnam, Nueva Zelanda y Colombia. Mientras que países como Qatar, Corea del Sur, Turquía, Serbia, Egipto, Iraq y Perú también presentan tasas de crecimiento sólidas, superiores al 30%. Además, se observa un crecimiento constante en países como Pakistán, Irán, India, Malasia, Emiratos Árabes y Argelia, todos con una tasa media de crecimiento anual superior al 20%.

Estos datos resaltan la dinámica global en la investigación en ciencias sociales sobre big data, con un variado interés y crecimiento en diferentes regiones del mundo, indicando creciente importancia de este campo a nivel internacional.

#### ***4.3.1 Big data en los distintos continentes***

Si agrupamos la información generadas por continentes y regiones, los principales contribuyentes asiáticos (excluida China, dado su dispar preponderancia) se sitúan en las regiones sudoriental y oriental: India (528 publicaciones), Corea (370), Japón (229) y Taiwán (209).

África en su conjunto registra 269 publicaciones, la mayoría de ellas producidas por países como Sudáfrica (103), Marruecos (32), Egipto (28), Nigeria (27) y Túnez (20).

América Latina produjo 309 artículos. Brasil está en la delantera con 139 publicaciones. Otros contribuyentes proceden de Colombia (43), México (42), Chile (29), Ecuador (28) y Argentina (12). Al analizar estos datos, es evidente que Brasil se destaca como el principal contribuyente de la región, con un número de 139 publicaciones, representando casi la mitad del total latinoamericano. Este fenómeno refleja la influencia y el liderazgo de Brasil en el ámbito de la investigación académica en América Latina.

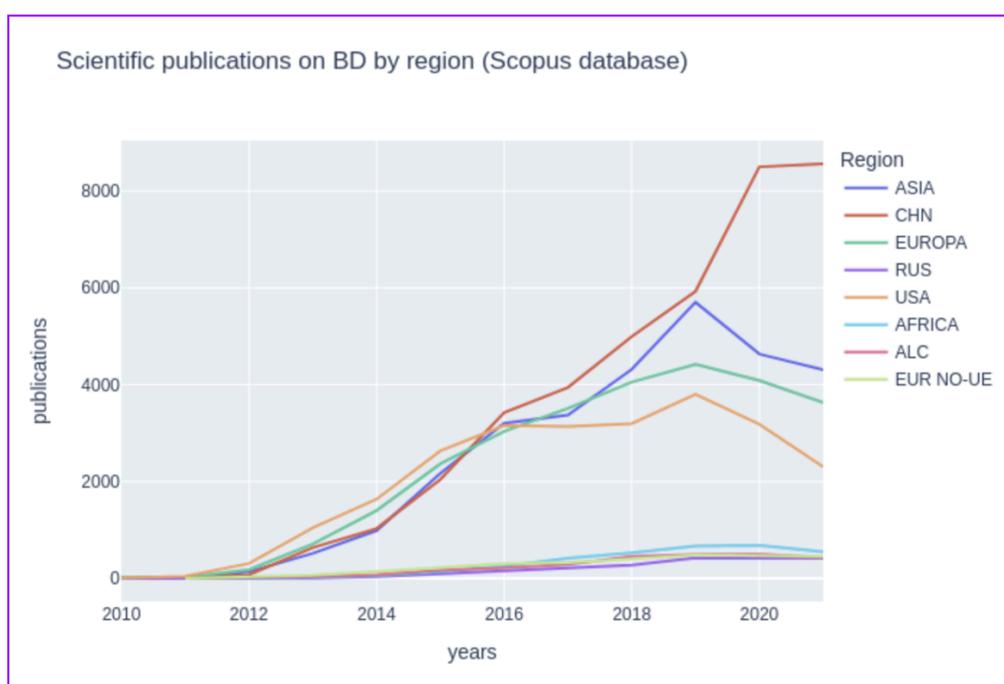
Esta tendencia de destacada contribución por parte de Brasil se confirma aún más al comparar estos resultados con los obtenidos en la base de datos Scielo. La coherencia entre diferentes fuentes de datos subraya y corrobora la posición preeminente de Brasil en la investigación académica de big data en el contexto latinoamericano. La calidad de la investigación y la experiencia de los investigadores brasileños en este campo han consolidado la posición de Brasil como un líder en la producción de conocimiento en ciencias sociales relacionadas con el campo.

La investigación en Europa Occidental está encabezada por Gran Bretaña (1.224), Alemania (560), España (482), Italia (447), Países Bajos (364) y Francia (274). Los países de Europa del Este registran un menor volumen de publicaciones que sus vecinos occidentales, pero mayores tasas de crecimiento como Rusia (239), Polonia (88) o Rumanía (88).

Por último, en Oriente Medio, los principales contribuyentes se encuentran en Arabia Saudí (94), Turquía (89), Oman (80), Pakistan (78) e Israel (75). Representando uno de los continentes con más baja contribución, en términos de textos académicos en Scopus.

La figura 26 abajo demuestra la distribución de publicaciones al recorrer de los años en las distintas regiones del globo. China (color rojo) fue separada del resto de Asia por su prolifera participación. La sigla ALC, color fucsia, representa publicaciones de América Latina e Caribe, región de interés prioritario de este estudio.

**Figura 26.** *Publicaciones sobre big data por año y por región.*



En lo que respecta a las publicaciones en América Latina y el Caribe, es evidente que Brasil sobresale como líder en este campo. México se sitúa en la segunda posición en términos de contribución, con un crecimiento constante en el grupo de países más destacados.

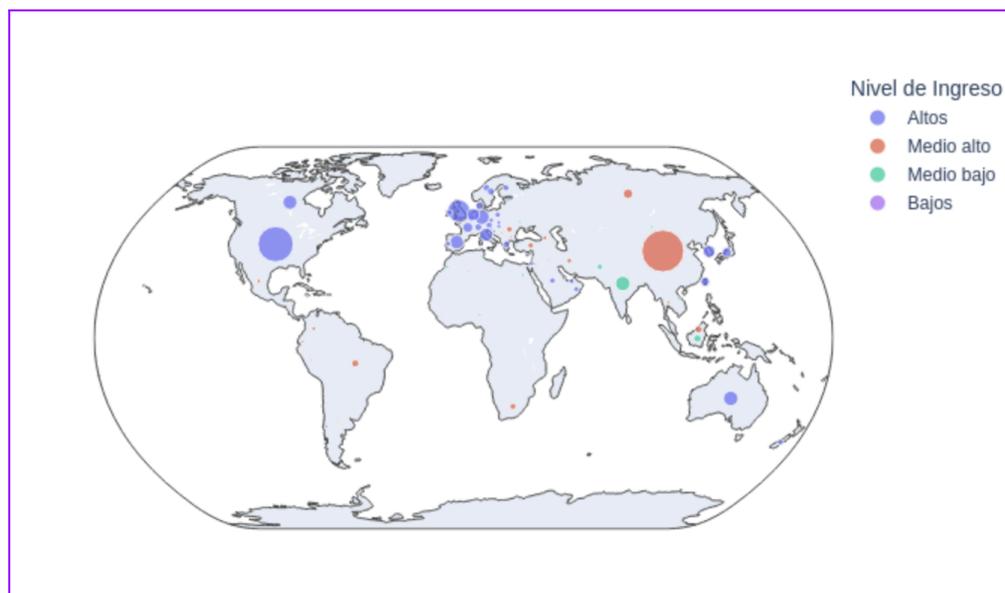
Sin embargo, es interesante observar que tanto Brasil como México experimentaron un descenso en sus publicaciones a partir del año 2020. En contraste, Colombia es un caso único en el continente que mantiene un crecimiento sostenido en su producción de investigación en este campo, y es el único país que sigue esta tendencia de crecimiento constante hasta la fecha. A pesar de su menor visibilidad inicial en el mapa de publicaciones, la publicación colombiana ha demostrado una resiliencia constante en su crecimiento académico en el área, lo que podría indicar un compromiso continuo con la investigación y la innovación en los estudios en big data. A continuación, se representa específicamente el crecimiento de las publicaciones en ciencias sociales en el continente latinoamericano:

**Figura 27.** *Publicaciones sobre big data en ciencias sociales en los países latinoamericano*



Volviendo al análisis de las publicaciones en Scopus y con la perspectiva global de los datos, se percibe que la mayoría de los textos indexados en esta librería proceden de países de renta alta (78%) y media-alta (40%). Las contribuciones de países de renta media-baja (7%) y baja (0,2%) son menos representativas, como se puede observar en la figura a continuación.

**Figura 28.** *Publicaciones sobre big data por país y grupo de ingresos*



Una atención a este hallazgo, nos puede conducir a algunas conclusiones e indicios relevantes. El primero es con respecto a la desigualdad en la contribución científica: la mayoría de los textos indexados en Scopus provienen de países de renta alta y media-alta, lo que indica una marcada desigualdad en la contribución científica global. Esto sugiere que los países con recursos económicos más sólidos tienen una presencia significativamente mayor en la investigación relacionada con big data en ciencias sociales, como también comprueba y sugiere, que los países latinoamericanos, mayoría proveniente de la parte más desigual y al sur del globo, experimenten menos participación de esta librería.

En segundo lugar, tal realidad configura un desafío para países en desarrollo: la baja representación de países de renta media-baja y baja en la producción científica puede indicar desafíos en el acceso a recursos y financiamiento para investigaciones en estas regiones. Estos países pueden enfrentar limitaciones en términos de infraestructura, capacitación y financiamiento para llevar a cabo investigaciones en el campo de big data en ciencias sociales, lo cual, como pudimos observar es no sólo necesario pero determinante para la

investigación y los contextos que estudien, utilicen o manipulen una estructura de procesamiento computacional de grandes datos.

Un tercer factor que implicaría el circuito y logística operacional del estudio en/para/con big data es la necesidad de colaboración internacional, que debido a la disparidad en la contribución científica, resalta su importancia en la investigación relacionada con los datos masivos. Los países de renta media-baja y baja podrían beneficiarse de alianzas estratégicas con países de renta-alta y media-alta para acceder a recursos y conocimientos adicionales en este campo. Sobre esto, esta investigación ha sido prueba de ello, durante su conducción, he logrado acceder a los circuitos y redes de colaboración en estudios y investigaciones aplicadas a big data en las ciencias sociales.

Con relación a este mismo aspecto, es sentido común el entendimiento que los países de renta-media alta; renta alta (los que por general son los detentores de las capacidades operacionales e infraestructuras en big data), ya disfrutan una enorme ventaja en cuanto al volumen de datos e información que todos generamos gratuitamente muchas veces, al ser usuario de sus servicios. Habría ahora que hacer que esta alianza sea equitativamente beneficiosa para ambas fuerzas.

Independientemente de la situación de desventaja en que puedan parecer estar los países del sur global, se destaca, por otro lado, el potencial de crecimiento que estas regiones tienen; es decir, la capacidad de crecer a medida que invierten en capacidades de investigación y desarrollo tecnológico. La distribución de la producción científica de big data en ciencias sociales por más que refleje desigualdades globales en términos de recursos y capacidades de investigación, destaca la importancia de abordar la brecha existente y de promover la colaboración internacional para impulsar el avance de la investigación en estas áreas en todo el mundo.

#### ***4.3.2 Autores que publican temáticas big data en las ciencias sociales***

Analizando la base de datos fue posible encontrar que apenas el 12% de estas publicaciones consisten en contribuciones individuales. La mayoría de ellas (75%) fueron producidas por equipos formados de 2 a 5 investigadores, el 12% por equipos de 6 a 10 y el 1% por equipos de más de 10 investigadores. Cabe señalar que 16 de estos últimos equipos fueron fruto de redes de colaboración de 50 a 100 investigadores y 8 de redes de más de 100 investigadores de 28 países. Como se puede percibir, la mayoría de las publicaciones son el resultado de

equipos de investigación, lo que sugiere que este campo requiere una variedad de habilidades y conocimientos que se combinan mejor en trabajos grupales que individuales. La existencia de equipos de investigación con más de 10 investigadores, incluidas las redes de colaboración de gran envergadura, destaca la complejidad y el alcance de la investigación en *big data*, sugiriendo una predisposición a abordar problemas complejos y esfuerzos colaborativos a gran escala.

Sin embargo, la mayoría de las redes de colaboración que están detrás de estas publicaciones son de ámbito nacional. Sólo el 19% son internacionales, de las cuales el 15% son binacionales, el 3% incluyen investigadores de 3 países, el 1% de 4 países y menos del 1% procedían de 5 o más países. A pesar de que existen redes de colaboración a nivel internacional, la mayoría de las colaboraciones son de ámbito nacional. Esto podría deberse a factores como la proximidad física, idioma común y recursos compartidos dentro de un país. Las colaboraciones binacionales son las más comunes, lo que puede apuntar a una cooperación cercana entre países vecinos.

En conjunto, el cuerpo científico que investiga sobre big data asciende a 168.463 investigadores en todo el mundo. Aunque hay investigadores de 158 países, 25 de ellos concentran el 89% de los investigadores, lo que refleja la desigual distribución mundial del conocimiento y el saber hacer en este campo. Las instituciones chinas (29%) y norteamericanas (19%), una vez más, son los principales centros de estos investigadores. La amplia distribución de investigadores, como es el caso, en 158 países, podría indicar que la temática big data viene creciendo en todo el mundo.

Sin embargo, la concentración de investigadores en unos pocos países principales sugiere que el conocimiento y los recursos están más desarrollados en estos lugares. Así, la concentración de investigadores en un pequeño número de países, especialmente en China y Estados Unidos, refleja una desigualdad significativa en la distribución global del conocimiento y la inversión en investigación en el campo big data. Esto también puede estar relacionado con la inversión en infraestructura y otros recursos en estos países, como se ha mencionado.

Los países europeos, por su parte, cuentan en conjunto con alrededor del 17%, la India tiene el 6% del total de los investigadores. Otros países como Rusia, Brasil, Indonesia o Malasia tienen entre 1.000 y 2.000 investigadores cada uno; Turquía, Marruecos, Polonia, Pakistán, Irán, Sudáfrica o Israel tienen entre 500 y 1.000 investigadores y México, Colombia, Egipto, Argentina, Vietnam, Chile o Ecuador entre 100 y 500 investigadores.

La gran mayoría de estos investigadores (77%) son colaboradores ocasionales, mientras que el 22,4% ha participado en 2 a 10 artículos. Sin embargo, el núcleo de este cuerpo científico, compuesto por autores con más de 10 publicaciones, sólo cuenta con unos 1.000 investigadores. Entre ellos, los 10 primeros autores tienen más de 50 publicaciones y están radicados principalmente en China, Estados Unidos, Reino Unido e India, pero también en Italia, Canadá, Australia, España, Arabia Saudí, Corea y Portugal.

**Tabla 8.** *Los 10 principales autores de las publicaciones en ciencias sociales sobre big data en el mundo*

Autor	País	Número de artículos
Bibri S.	Noruega	24
Li Q.	China	20
Liu Y.	China	19
Batty M.	Gran Bretaña	19
Li Z.	Estados Unidos	15
Ye X.	Estados Unidos	15
Pettit C.	Australia	15
Jiang B.	Suecia	14
Leung C.	Canadá	14
Mosavi A.	Alemania	13

Con respecto al nivel de renta del país, los principales autores de los países de renta media-alta son principalmente chinos, sudafricanos y colombianos, con entre 35 y 50 publicaciones. Los principales investigadores de los países de renta media-baja pertenecen sobre todo a instituciones de India y Marruecos. Estos investigadores han producido entre 18 y 35 artículos. Por último, los investigadores de países de renta baja proceden principalmente de África, Oriente Medio y Nepal. Sin embargo, su productividad sigue siendo muy baja: 2 artículos por autor, salvo en el caso de Sun (12) y Maharjan (4) como puede observarse a continuación.

**Tabla 9.** *Los 5 principales contribuyentes por nivel de renta en el país*

Nivel de Ingreso	Autor	País	Número de publicaciones
Altos	Bibri S.	Noruega	24
Altos	Batty M.	Gran Bretaña	19
Altos	Pettit C.	Australia	15
Altos	Ye X.	Estados Unidos	15
Altos	Li Z.	Estados Unidos	15
Medio alto	Li Q.	China	20

Medio alto	Liu Y.	China	19
Medio alto	Tu W.	China	11
Medio alto	Niu L.	China	11
Medio alto	Pei T.	China	10
Medio bajo	Harold Robinson Y.	India	6
Medio bajo	Hasan R.	Pakistán	6
Medio bajo	Sun Z.	Papúa Nueva Guinea	5
Medio bajo	Shukla D.	India	5
Medio bajo	Hassanien A.	Egipto	5
Bajos	Yélérou T.	Burkina Faso	2
Bajos	Tall H.	Burkina Faso	2
Bajos	An Y.	Corea del Norte	2
Bajos	Bekunda M.	Tanzania	1
Bajos	Dandjinou M.	Burkina Faso	1

En cuanto a los autores en América latina trabajando en big data<sup>40</sup>, se observa una representación de autores provenientes de distintos países del continente.

**Tabla 10.** Principales autores por país que investigan sobre big data en América Latina

Región	Autor	País	Nº de publicaciones
ALC	Maçada A.	BRA	5
ALC	Câmara G.	BRA	5
ALC	Vinhas L.	BRA	4
ALC	Hoppen N.	BRA	3
ALC	Cacho N.	BRA	3
ALC	Munizaga M.	CHL	4
ALC	Villegas-Ch W.	ECU	4
ALC	Cervantes O.	MEX	3
ALC	Lu K.	PRI	3
ALC	Ochoa D.	ECU	2

Como se puede inferir a partir de los datos, la investigación sobre big data en ciencias sociales es un campo en crecimiento en América Latina, e incluso incipiente si se compara con otras regiones. De los 46 países que conforman América Latina y Caribe, apenas 5 han contribuido con publicaciones mencionadas en la base de datos Scopus. La poca diversidad

<sup>40</sup>Para más información acerca de la distribución de los autores más influyentes de las demás partes del globo, revisar el anexo de esta investigación.

geográfica resalta la importancia de la colaboración en toda la región. Por otro lado, Brasil emerge como un líder en la producción de investigación sobre big data en las ciencias sociales latinoamericanas, con varios autores y un alto número de publicaciones. Esto sugiere un interés significativo y una capacidad de investigación en este campo en el país.

El crecimiento constante de la investigación sobre big data en las ciencias sociales latinoamericanas, así como las tasas de crecimiento en países como México, Ecuador y Chile, indican un interés continuo y un desarrollo en este campo, representando un futuro de potencial para la investigación en la región<sup>41</sup>.

### ***4.3.3 Instituciones más relevantes para la difusión de temas big data en las Ciencias Sociales***

De las 20 principales instituciones que acogen a estos investigadores, 18 son institutos de investigación y universidades chinas (Universidad de Nanjing, Universidad de Tsinghua, Universidad de Wuhan, Universidad Nacional, Universidad de Correos y Telecomunicaciones de Pekín, Universidad de Pekín, Universidad de Beihang, Universidad de la Academia China de Ciencias, Universidad Jiao Tong de Shanghai, Universidad de Zhejiang, entre otras). En promedio, cada una de estas universidades ha publicado unos 1.500 artículos sobre big data. En el puesto 12, con unas 1.000 publicaciones, encontramos a la Universidad de Delhi, de la India, y en el 16, al Instituto Tecnológico de Massachusetts (MIT), de EE.UU., con aproximadamente 800 artículos. Desglosadas por nivel de renta de los países, las diez instituciones chinas mencionadas ostentan un liderazgo indiscutible entre los países de renta media-alta.

Las diez primeras instituciones de los países de renta más baja se encuentran en India, Indonesia y Marruecos. En promedio, han publicado unos 200 trabajos, salvo la Universidad de Delhi, que tiene alrededor de 1.000 publicaciones. Entre los países de renta alta figuran otras instituciones norteamericanas (MIT, Purdue, Michigan, California del Sur, Washington, Minnesota y Virginia Tech), la Universidad de Tokio (Japón), el Politécnico de Milán (Italia) y la Universidad de Melbourne. Por término medio, cada una de estas instituciones tiene aproximadamente 500 publicaciones. Por último, entre los países de renta baja, encontramos principalmente universidades africanas de Ruanda, Uganda, Etiopía, Níger y Madagascar,

---

<sup>41</sup>Para conocer con mayor profundidad las redes de colaboración internacional de los autores de America Latina, acceder al siguiente enlace interactivo: <http://viz.divergencelab.org/bigdataSoc/netBDSocAutReg.html>.

pero también de Nepal y Siria. Estas instituciones han publicado entre 3 y 15 artículos cada una.

**Tabla 11.** Las 5 principales instituciones que investigan sobre big data por nivel de ingresos y país

Nivel de Ingreso	Institución	País	Publicaciones
Altos	Hong Kong Polytechnic University	Hong Kong	42
Altos	UNSW Sydney	Australia	33
Altos	University College London	Gran Bretaña	32
Altos	Arizona State University	Estados Unidos	24
Altos	Technical University of Munich	Alemania	24
Medio alto	University of Chinese Academy of Sciences	China	65
Medio alto	Peking University	China	56
Medio alto	Tsinghua University	China	46
Medio alto	Wuhan University	China	44
Medio alto	Beihang University	China	36
Medio bajo	Vellore Institute of Technology	India	14
Medio bajo	Institut Teknologi Bandung	Indonesia	9
Medio bajo	Universitas Gadjah Mada	Indonesia	8
Medio bajo	Bina Nusantara University	Indonesia	8
Medio bajo	Anna University	Indonesia	7
Bajos	State Academy of Sciences, Institute of Development Strategy of Science and Technology	Corea del Norte	2
Bajos	Université Nazi Boni (UNB)	Burkina Faso	2
Bajos	University of Abomey-Calavi	Benin	1
Bajos	Institut de Mathematiques et de Sciences Physiques University of Abomey-Calavi	Benin	1
Bajos	Institute of Engineering Pulchowk	Nepal	1

Por otro lado, en América Latina percibimos las siguientes distribuciones por institución de procedencia del autor. Para ver la tabla completa por cada región del globo, revisar el anexo.

Como se puede notar la USP (Universidad de São Paulo) es la número uno en publicaciones en ciencias sociales y big data. Cabe mencionar que las universidades brasileñas que aparecen rankeadas son instituciones públicas, lo que remarca aún más la comprensión de que el conocimiento sobre la temática, por lo menos en América Latina, sobrepasa el ámbito público y tal vez gubernamental.

**Tabla 12.** *Las 5 principales instituciones que investigan sobre big data en América Latina*

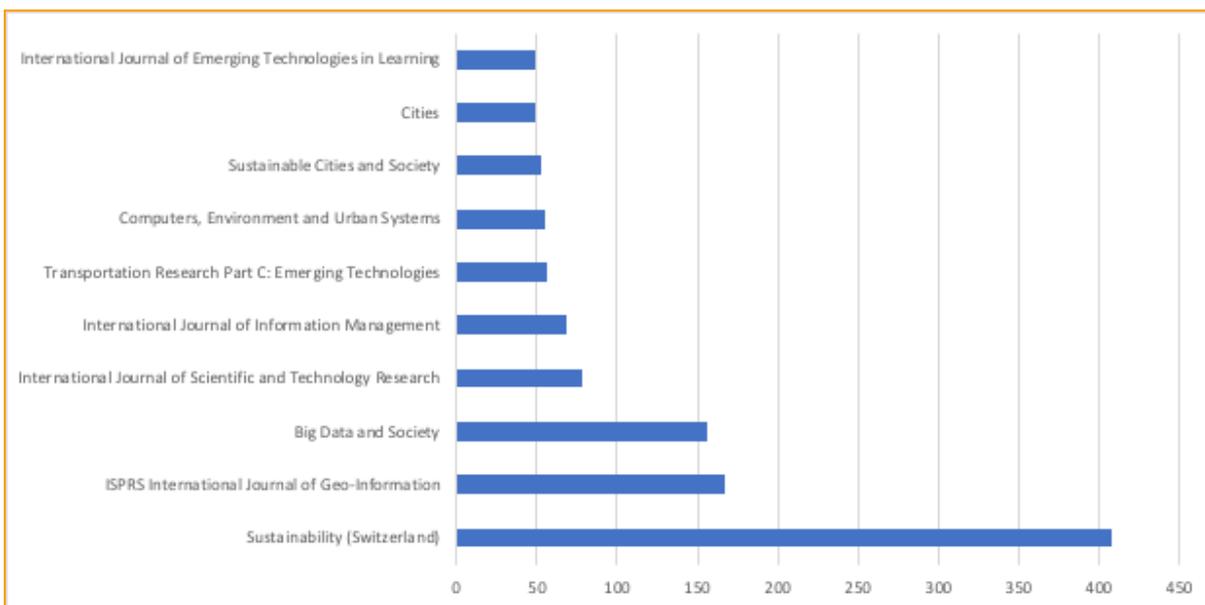
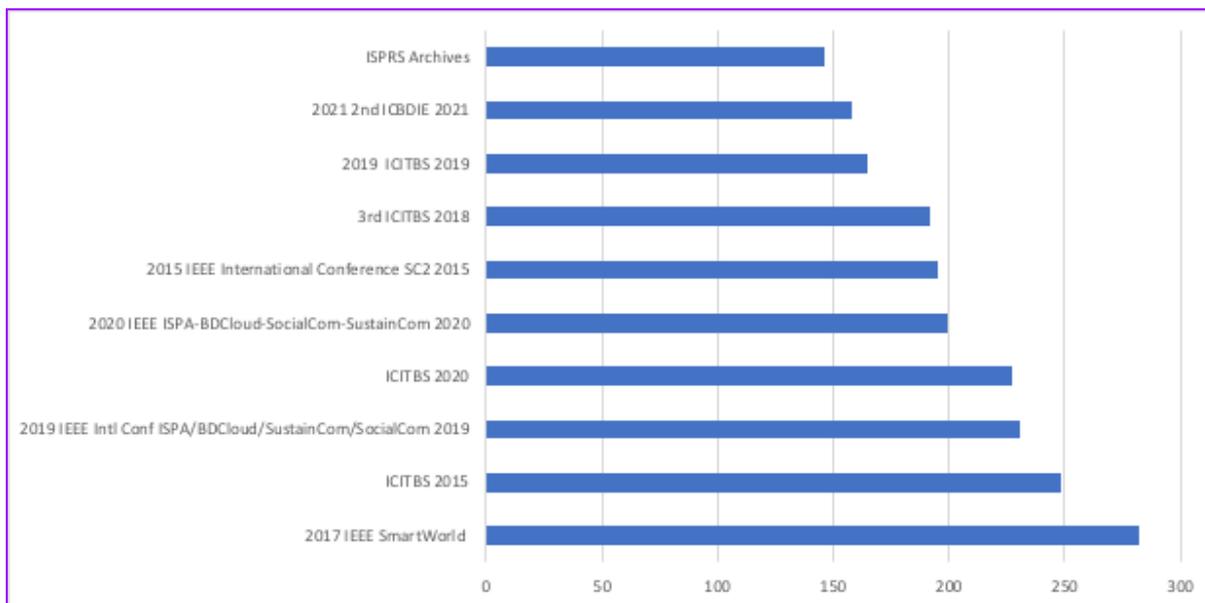
Región	Institución	País	Publicaciones
Brasil	Universidade de São Paulo	BRA	20
Brasil	Universidade Federal do Rio Grande do Sul	BRA	7
Brasil	Universidade Estadual Paulista "Júlio de Mesquita Filho"	BRA	6
Brasil	Instituto Nacional de Pesquisas Espaciais	BRA	6
Brasil	Universidade Federal de Santa Catarina	BRA	6
América Latina	Universidad Nacional Autónoma de México	MEX	6
América Latina	Tecnológico de Monterrey	MEX	4
América Latina	Pontificia Universidad Católica de Chile	CHL	4
América Latina	Pontificia Universidad Católica del Perú	PER	4
América Latina	Universidad de las Américas - Ecuador	ECU	4

*Nota:* El color azul representa instituciones públicas; el rosado, las privadas

#### **4.3.4 Revistas y conferencias relevantes sobre big data en las Ciencias Sociales**

Para completar esta visión general de la productividad científica y los indicadores de rendimiento de la investigación global sobre big data, sería valioso identificar las principales conferencias y revistas científicas que presentan y publican los resultados de estos estudios, ilustrado a continuación.

**Figura 29.** Principales actas de congresos y revistas científicas que publican investigaciones sobre big data en ciencias sociales (a + b)



Nota: El color morado representa letra “a”; el naranja, “b”

Las actas publicadas en la serie *Lecture Notes in Computer Science* son el principal ágora para este tipo de investigación. Otras publicaciones importantes para artículos de conferencias son las *ACM International Conference Proceeding*, las *Communications in Computer and Information Science*, las *Advances in Intelligent Systems and Computing* y las

*conferencias* anuales del IEEE sobre el tema de los macrodatos. Las 20 primeras de estas conferencias representan el 34% de la producción científica total sobre este tema.

Como ya se ha indicado, los artículos de revistas son menos numerosos y están distribuidos de forma más uniforme, cuando hablamos de la tendencia general de las publicaciones en todas las áreas del conocimiento. *IEEE Access* es la revista líder con más de 470 artículos publicados, seguida de *Future Generation Computer Systems*, *Cluster Computing*, *International Journal of Innovative Technology and Exploring Engineering*, *International Journal of Recent Technology and Engineering* y *Journal of Big Data* con una media de 200 artículos cada una. A diferencia de las publicaciones en congresos, las 20 primeras revistas sólo representan el 14% del total de este tipo de publicaciones.

En cuanto a las estrategias de publicación de los autores pertenecientes a los distintos países, se encuentran salidas diferentes. Por ejemplo, los investigadores de países de renta alta y media publican en las mejores revistas, como *IEEE Access*, *Future Generation Computer System o Sustainability*. Sin embargo, los 7 últimos periódicos de los diez primeros, para cada categoría de ingresos son diferentes. En los países de renta media-baja, sólo *IEEE Access* ocupa una posición en la lista de las diez principales revistas, mientras que en los países de renta baja, el *Journal of Big Data* es la principal publicación utilizada para presentar los resultados de este tipo de investigación, lo que me hace traer a la memoria mi primer acercamiento a esta bibliografía, pues fue de donde provinieron los primeros textos que leí, de donde surge la gran parte de mi inspiración y la idea de construcción de las categorías de análisis que utilizo para entender los datos y fabricar el conocimiento acá presentado.

El resto de revistas utilizadas por los investigadores de países de renta media-baja y baja para publicar sus resultados son distintas en cuanto a (Tabla 13) que, a pesar de la disponibilidad de algunas revistas que articulan las investigaciones y debates de distintas latitudes, las comunidades científicas de diferentes regiones geoeconómicas se reproducen por caminos diferentes.

**Tabla 13.** Principales revistas científicas que publican artículos relacionados con big data en ciencias sociales por grupo de países.

Grupo de países por nivel de ingreso	Revista	Publicaciones
Altos	Sustainability (Switzerland)	255
Altos	Big Data and Society	155
Altos	ISPRS International Journal of Geo-Information	79
Altos	International Journal of Information Management	55
Altos	Computers, Environment and Urban Systems	43
Medio alto	Sustainability (Switzerland)	137
Medio alto	ISPRS International Journal of Geo-Information	88
Medio alto	International Journal of Emerging Technologies in Learning	40
Medio alto	Journal of Transportation Systems Engineering and Information Technology	35
Medio alto	Dili Xuebao/Acta Geographica Sinica	33
Medio bajo	International Journal of Scientific and Technology Research	69
Medio bajo	Library Philosophy and Practice	14
Medio bajo	Sustainability (Switzerland)	14
Medio bajo	Webology	10
Medio bajo	Information Processing and Management	8
Bajos	American Behavioral Scientist	1
Bajos	Geocarto International	1
Bajos	International Journal of Scientific and Technology Research	1
Bajos	Internet Policy Review	1
Bajos	Journal of Engineering Education Transformations	1

Con respecto a América Latina específicamente, la tabla proporciona información sobre las principales revistas científicas que han servido como plataforma para la publicación de artículos relacionados con big data en el ámbito de las ciencias sociales en América Latina y el Caribe. A continuación, se realiza una descripción detallada de los datos.

A partir de ellos, se puede concluir que estas revistas juegan un papel fundamental en la difusión de la investigación sobre big data en ciencias sociales en América Latina y el Caribe. La variedad de revistas y el número de publicaciones sugieren un interés creciente en esta área de estudio en la región. Además, la inclusión de revistas interdisciplinarias, como

*Sustainability Switzerland* y *Research in Transportation Economics*, destaca la naturaleza diversa de las investigaciones sobre big data que abarcan temas más allá de las ciencias sociales tradicionales. Estas revistas brindan una plataforma valiosa para académicos e investigadores que desean contribuir al avance de esta área de conocimiento en la región.

**Tabla 14.** Principales revistas científicas que publican artículos relacionados con big data en ciencias sociales por región.

Región	Journal	Publicaciones
ALC	Perspectivas em Ciencia da Informacao	6
ALC	Ciencia da Informacao	5
ALC	Sustainability (Switzerland)	5
ALC	Research in Transportation Economics	3
ALC	Revista Brasileira de Políticas Públicas	3

Como se puede ver, el campo de "ciencias de la información" se destaca como una área prominente en la tabla de revistas científicas que publican artículos relacionados con big data. Este reconocimiento se basa en la presencia significativa de revistas escritas en portugués y afiliadas a instituciones brasileñas en la lista. La alta representación de revistas como *Perspectivas em Ciência da Informação* y *Ciência da Informação*, ambas afiliadas a instituciones en Brasil, resalta la influencia y el interés de la comunidad académica brasileña en este tema.

Aunque de origen brasileño, estas revistas también tienen una proyección internacional, lo que refuerza su importancia. Además, la inclusión de revistas extranjeras como *Sustainability* y *Research in Transportation Economics* destaca la colaboración internacional y la interdisciplinariedad en la exploración de big data, lo que amplía aún más la relevancia de este campo en la región.

Sin embargo, los trabajos presentados en congresos muestran un panorama diferente. *Lecture Notes in Computer Science*, *ACM International Conference* y las conferencias del *IEEE* son las principales ágoras para los científicos de países de todas las niveles de renta. Otras publicaciones basadas en conferencias, como *Advances in Intelligent Systems and Computing* y *Communications in Computer and Information Sciences*, parecen ser más frecuentadas por investigadores de países de renta alta y media-alta. Del mismo modo, conferencias como *ICSITech 2017*, *Confluence 2016* e *ICSITech 2017* atrajeron a más investigadores de países de

renta media-baja, mientras que *ICITBS 2018*, *SEBD 2015* y *OBD 2015* captaron la atención de países de renta baja (tabla 15).

**Tabla 15.** Principales conferencias científicas en las que se han presentado trabajos relacionados con los big data en ciencias sociales por grupo de países

Grupo de país por nivel de ingreso	Conferencia	Publicaciones
Altos	2017 IEEE SmartWorld	207
Altos	2015 IEEE SC2 2015	97
Altos	ISPRS Archives	66
Altos	2016 IEEE SustainCom 2016	50
Altos	2020 IEEE ISPA-BDCloud-SocialCom-SustainCom	49
Medio alto	2015 ICITBS 2015	246
Medio alto	2020 ICITBS	221
Medio alto	2019 IEEE ISPA/BDCloud/SustainCom/SocialCom	195
Medio alto	3rd ICITBS 2018	188
Medio alto	2019 ICITBS	160
Medio bajo	2015 ICSITech	49
Medio bajo	INDIACom 2016	36
Medio bajo	2016 3rd ICBDESC	16
Medio bajo	2017 IEEE SmartWorld	10
Medio bajo	ISPRS Archives	10
Bajos	IBASE-BF 2020	5
Bajos	2019 IEEE PuneCon 2019	1
Bajos	2020 INCET	1
Bajos	2015 ICITBS	1

Ya en América Latina, la de mayor expresividad es *ISPRS archives*<sup>42</sup> (*International Society for Photogrammetry and Remote Sensing*). Como se puede percibir en la tabla abajo. En segundo lugar está *IEEE SmartWorld*<sup>43</sup>, seguido por *Proceedings of the Brazilian Symposium*

<sup>42</sup> <https://www.isprs.org/proceedings/xxxv/congress/comm4/comm4.aspx> accedido en setiembre de 2022

<sup>43</sup> 2017 IEEE SmartWorld Ubiquitous Intelligence and Computing, Advanced and Trusted Computed, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People and Smart City Innov - Universities and research institutions in United States - <https://www.scimagojr.com/journalsearch.php?q=21100868894&tip=sid&clean=0> accedido en septiembre de 2022.

on *GeoInformatics*<sup>44</sup>. Como se puede notar hay una predominancia muy fuerte de trabajos sobre la temática satelital, de las 5 principales conferencias científicas que han presentado trabajos relacionados a big data, 3 de ellos son de enfoque de datos informática, y los otros dos en congresos sobre Tecnología (tabla 16).

**Tabla 16.** *Principales conferencias científicas en las que se han presentado trabajos relacionados con los big data en ciencias sociales por regiones*

Región	Conferencia	Publicaciones
ALC	ISPRS Archives	7
ALC	2017 IEEE SmartWorld	4
ALC	Proceedings of the Brazilian Symposium on GeoInformatics	4
ALC	Proceedings of the LACCEI international Multi-conference for Engineering, Education and Technology	4
ALC	2020 CONIITI	2

La relación entre big data y datos geo informacionales (geo-espaciales o geográficos) es una convergencia frecuente y se ha demostrado con un potencial importante para los estudios poblacionales y de infraestructura social, sobre todo en la actualidad. La emergencia de las tecnologías big data y el procesamiento computacional de datos ha sido especialmente relevante en el ámbito de la recopilación de información geoespacial. Hoy en día, una variedad de sensores y dispositivos, como satélites, drones, GPS, sensores de IoT (Internet de las cosas) y aplicaciones móviles, generan constantemente grandes cantidades de datos geoespaciales. Estos datos incluyen información sobre ubicaciones, coordenadas geográficas, mapas de calor, imágenes aéreas y mucho más. Con la potencialidad de manipular no solamente grandes cantidades de datos, además de otros tipos de formatos de datos, estamos experimentando, una expansión de la capacidad de investigar e incidir en campos del conocimiento de gran escala, y mucho más integrado con diversas disciplinas y producciones de conocimiento.

A menudo, los datos geoespaciales son inherentemente complejos debido a su naturaleza multidimensional. Big data proporciona la capacidad de analizar y procesar estas vastas cantidades de datos para descubrir patrones espaciales, tendencias geográficas y relaciones que antes eran difíciles de identificar. Esto es crucial en disciplinas como la geografía, la

<sup>44</sup><http://www3.cptec.inpe.br/dimnt/2022/09/28/geoinfo22-xxiii-brazilian-symposium-on-geoinformatics-sao-jose-dos-campos-brazil-november-28-30-2022/> accedido en setiembre de 2022

planificación urbana, la gestión de recursos naturales y la epidemiología, donde comprender la dimensión espacial es fundamental (Liu et al., 2016).

Otro factor que explica y justifica la íntima relación entre datos masivos, estudios espaciales y ciencias sociales (geografía social, para algunos) es el de predicción y toma de decisiones: La combinación de big data y datos geoespaciales permite la creación de modelos predictivos avanzados. Por ejemplo, en la gestión de desastres naturales, se pueden utilizar datos geoespaciales para predecir la trayectoria de un huracán o la propagación de un incendio forestal, lo que facilita una toma de decisiones más rápida y precisa.

De hecho, los datos geoespaciales impulsados por los datos masivos son la base de muchas aplicaciones que usamos en la vida cotidiana. Esto incluye servicios de mapas en línea, navegación GPS, aplicaciones de viajes compartidos como *Uber*, *ifood*, *Waze*, *Yo me cuido*, *Clima tiempo*, entre tantas otras. Estas aplicaciones utilizan datos geoespaciales para proporcionar información en tiempo real y facilitar la toma de decisiones no solamente de un ciudadano común, si no que influyen en las infinitas y aparentemente micro decisiones individuales que repercuten en el bien común, como la gestión del tráfico y operaciones de seguridad en situaciones de festividades o calamidades públicas, por ejemplo.

Este es el caso de una de las entrevista más interesantes que realicé a un trío de autores de unos de los artículos publicado por Scielo: los juegos olímpicos de Río de Janeiro, en 2014, en Brasil, originó y perpetuo uno de los centro de operación de comando y gestión del tránsito de unas de las ciudades más grandes del mundo. Seguramente el más tecnológico y eficiente de latinoamérica, gracias a una colaboración pública privada, entre el Estado de Río de Janeiro y una de las aplicaciones de movilidad urbana más importantes e influyentes del mundo.

Definitivamente la existencia de los datos masivos incrementó el estudio, manipulación y procesamiento de imágenes georreferenciadas de todo el mundo en nombre del beneficio social, del medio ambiente y el planeta Tierra: la monitorización ambiental a través de sensores geoespaciales es fundamental para comprender y mitigar problemas ambientales como la deforestación, el cambio climático, la gestión de recursos hídricos y la conservación de la biodiversidad. Permitir el análisis de datos geoespaciales a gran escala, resulta en una mejor comprensión de los patrones ambientales y la toma de decisiones informadas global e internacionalmente.

Yo misma, cuando ejercía mis trabajos como etnógrafa en ciencias de datos, era responsable por la conducción de la investigación cualitativa en un proyecto Fondef, creado con el objetivo de facilitar, entre otras cosas, el procesamiento de una masiva cantidad de datos satelitales que diversos sectores gubernamentales de Chile requerían.

Aunque este dato no puede ser tomado como una “evidencia que comprueba un hecho en función del otro”, mi propia trayectoria profesional corrobora esta tendencia, lo que creo, no es coincidencia. El Fondef *IDEA I+D - SAMSARA*<sup>45</sup>, que justamente trataba sobre imágenes satelitales, medición y formación de información, vía el estudio de las fotos y mediciones remotas vía satélite, consistía en una investigación cualitativa sobre la opinión de los más diversos ministerios y servidores públicos involucrados en la utilización y procesamiento de información satelital en Chile.

#### 4.3.5 Citas más relevantes

El número de citas de un artículo denota su relevancia o influencia en el campo. El conjunto de datos utilizado para este estudio muestra que la mayor parte de esta producción científica (42%) no tiene ninguna cita. Esto significa que sus resultados y conocimientos aún no han encontrado eco en sus respectivas comunidades científicas. Otro 38% tiene entre una y cinco citas, un 8% entre 5 y 10 y un 9% hasta 50 citas. El 1,8% tiene más de 50 citas, lo que representa el 41% del total de citas en este campo. Entre estos últimos, unos 30 artículos han sido citados más de 500 veces y unos 10 artículos más de 1.000 veces. Sin lugar a dudas, estos son los artículos seminales en el campo (Tabla 17).

**Tabla 17.** Los 10 artículos más citados con la palabra clave “big data” en ciencias sociales

Título	Revista o Conferencia	Autores	Año	País	Citas
Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon	Information Communication and Society	Boyd D., Crawford K.	2012	USA	2728
Beyond the Hype: Big data Concepts, Methods, and Analytics	International Journal of Information Management	Gandomi A., Haider M.	2015	CAN	1928

<sup>45</sup>Fondef IDEA I+D, buscará durante dos años, desarrollar un sistema automático de monitoreo de imágenes satelitales de humedales urbanos, turberas de Chiloé y el bosque y matorral esclerófilo de la Región Metropolitana para mejorar la cobertura espacial y temporal de fiscalización de la Superintendencia del Medio Ambiente. Para acceder y descargar el informe final de la investigación cualitativa del proyecto SAMSARA: <https://goblabs.uai.cl/proyecto-samsara/> y acá los resultados de la encuesta interactiva: <https://lookerstudio.google.com/u/0/reporting/6b707b45-f2f8-42ed-a592-50fcaec965fe/page/f5N6C>

The Real-time city? Big Data and Smart Urbanism	GeoJournal	Kitchin R.	2014	IRL	1254
Big Data, New Epistemologies and Paradigm Shifts	Big Data and Society	Kitchin R.	2014	IRL	932
Big other: Surveillance Capitalism and the Prospects of an Information Civilization	Journal of Information Technology	Zuboff S.	2015	USA	846
Service Research Priorities in a Rapidly Changing Context	Journal of Service Research	Ostrom A.L., Parasuraman A., Bowen D.E., Patrício L., Voss C.A.	2015	USA	791
Datafication, Dataism and Dataveillance: Big data Between Scientific Paradigm and Ideology	Surveillance and Society	van Dijck J.	2014	NLD	609
The Role of Big Data in Smart City	International Journal of Information Management	Hashem I.A.T., Chang V., Anuar N.B., Adewole K., Yaqoob I., Gani A., Ahmed E., Chiroma H.	2016	MYS	545
Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data	Journal of Communication	Colleoni E., Rozza A., Arvidsson A.	2014	ITA	531
Big Data, Smart Cities and City Planning	Dialogues in Human Geography	Batty M.	2013	GBR	528

Cabe señalar que, a pesar del liderazgo de la producción china en este campo, la influencia de los investigadores afiliados a instituciones estadounidenses en estas publicaciones seminales es más dominante. Investigadores alemanes, británicos y españoles también han contribuido a esta literatura básica. Además, algunas de estas publicaciones revelan la existencia de redes de colaboración entre los autores chinos, británicos y norteamericanos de estos artículos. Si nos fijamos en los 20 autores con más de una publicación, resulta evidente que la mitad de ellos son investigadores afiliados a instituciones norteamericanas. La otra mitad está formada por investigadores afiliados a Malasia, China, Reino Unido, Sudáfrica y Georgia (Tabla 18).

**Tabla 18.** *Los 10 autores más citados en ciencias sociales sobre big data*

Autor	País	Afiliación	Citados por
Boyd D.	USA	Microsoft Research Cambridge	2879
Kitchin R.	IRL	Maynooth University	2553
Haider M.	CAN	Ted Rogers School of Management	1928
Batty M.	GBR	University College London	1118
Anuar N.	MYS	Universiti Malaya	778
Krogstie J.	NOR	Norges Teknisk-Naturvitenskapelige Universitet	759
Floridi L.	GBR	Alan Turing Institute	757
Andrejevic M.	AUS	Monash University	616
van Dijck J.	NLD	Universiteit Utrecht	609
Kang C.	CHN	Wuhan University	527

Según nuestros datos, Guizani parece ser el autor más influyente en este campo, con 25 publicaciones y más de 2.600 citas. Sin embargo, el 85% de estas citas se refieren a uno sólo de sus artículos: *Internet de los objetos: A Survey on Enabling Technologies, Protocols, and Applications*. Otros autores, como los del equipo español que están detrás del artículo principal (Tabla 17), han sido excluidos de esta lista, ya que sólo han hecho una contribución a este ámbito de investigación. La mayoría de estas investigaciones son interdisciplinarias, sugiriendo que muchas citas presentes en estos artículos podrían no estar relacionadas con big data y se originan en otras comunidades científicas.

Es preciso decir que el conjunto de datos utilizado para este estudio no contiene información suficiente para medir únicamente las citas dentro del mismo corpus. Esta cuestión podría abordarse en futuras investigaciones.

**Tabla 19.** *Los 5 autores más citados sobre big data por grupo de ingresos país*

Autor	Nivel de ingreso	País	Afiliación	Nº Citas
Boyd D.	Altos	USA/GBR	Microsoft Research Cambridge	2879
Kitchin R.	Altos	IRL	Maynooth University	2553
Haider M.	Altos	CAN	Ted Rogers School of Management	1928
Batty M.	Altos	GBR	University College London	1118

Krogstie J.	Altos	NOR	Norges Teknisk-Naturvitenskapelige Universitet	759
Anuar N.	Medio alto	MYS	Universiti Malaya	778
Kang C.	Medio alto	CHN	Wuhan University	527
Liu X.	Medio alto	CHN	Sun Yat-Sen University	424
Pei T.	Medio alto	CHN	Institute of Geographical Sciences and Natural Resources Research Chinese Academy of Sciences	347
Wang S.	Medio alto	CHN	Chinese Academy of Sciences	296
Sun Z.	Medio bajo	PNG	Papua New Guinea University of Technology	113
Alam M.	Medio bajo	IND	Jamia Millia Islamia	81
Bhattacharyya S.	Medio bajo	IND	Indian Institute of Management Nagpur	75
Kar A.	Medio bajo	IND	Indian Institute of Technology Delhi	69
Jagadeesh M.	Medio bajo	IND	Accenture, Global Information Management Services, India	69
Bekunda M.	Bajos	TZA	International Institute of Tropical Agriculture (IITA)	3
Zakari I.	Bajos	NER	Universite Abdou Moumouni	2

Con base en la tabla, si dividimos estos datos según las categorías de renta de los países, comprobamos que los diez autores más influyentes en los países de renta alta son todos investigadores afiliados a Norteamérica. Por término medio, cada autor tiene aproximadamente 10 publicaciones y entre 1.700 y 2.600 citas. En los países de renta media-alta, los investigadores afiliados a Malasia y China son los más influyentes; de media, tienen más publicaciones que sus homólogos norteamericanos (19) y entre 1.300 y 2.300 citas cada uno. Los 10 investigadores más influyentes de ingresos medios bajos proceden en su mayoría de la India. Por término medio, han publicado 14 artículos y han sido citados entre 190 y 1.000 veces cada uno. Por último, entre los investigadores de países de renta baja, Sun, afiliado a la PNG, University of Technology de Papúa Nueva Guinea, es el autor más prolífico, con unos 14 trabajos y más de 90 citas. Otros investigadores de estos países son los de Siria, Etiopía, Yemen, Benín, Madagascar y Nepal. Todos ellos tienen alrededor de 2 publicaciones, con entre 3 y 9 citas cada uno.

A continuación, se resume la información sobre los autores más destacados en el ámbito de las publicaciones académicas en ciencias sociales en América Latina.

**Tabla 20.** *Los 5 autores de ciencias sociales más citados sobre big data en América Latina*

Autor	Región	País	Afiliación	Citado por
Evsukoff A.	ALC	BRA	Universidade Federal do Rio de Janeiro	246
Bonilla S.	ALC	BRA	Universidade Paulista	183
Tavares E.	ALC	BRA	Universidade Federal do Rio de Janeiro	167
Estrella J.	ALC	BRA	Universidade de São Paulo	62
Telles R.	ALC	BRA	Universidade Paulista	54
Munizaga M.	ALC	CHL	Instituto Sistemas Complejos de Ingeniería	53
Ríos S.	ALC	CHL	Universidad de Chile	34
Cervantes O.	ALC	MEX	Universidad de las Américas Puebla	29
Forcael E.	ALC	CHL	Universidad del Bio Bio	27
Ochoa-Zezzatti A.	ALC	MEX	Universidad Autónoma de Ciudad Juárez	26

Los autores más prolíficos en esta selección de publicaciones en ciencias sociales de América Latina provienen principalmente de Brasil, lo que confirma una fuerte presencia académica en esta región. Los tres primeros autores son de la Universidad Federal de Río de Janeiro y la Universidad Paulista, ambas instituciones públicas brasileñas. Seguido, destacan autores de Chile y México, que a lo largo de dos años vienen manteniendo activas sus investigaciones en ciencias sociales en la región.

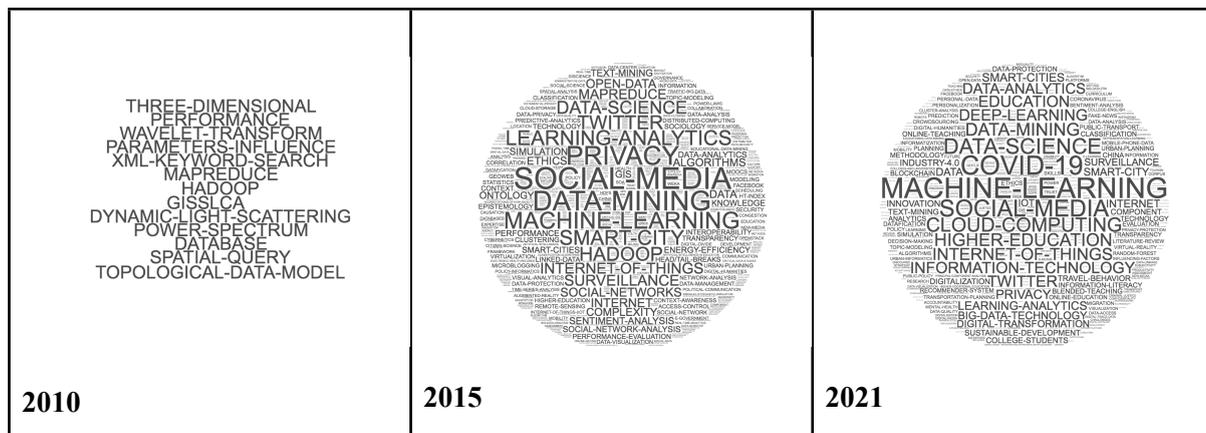
#### ***4.3.6 Tendencias de las palabras clave a lo largo del tiempo en las ciencia sociales mundiales***

Más allá de la clasificación por áreas de investigación, las palabras clave utilizadas para indexar estas publicaciones nos ofrecen una imagen más precisa de los principales temas estudiados dentro de la literatura sobre big data. También revelan las tendencias predominantes y la evolución de este tema de investigación a lo largo del tiempo. Las nubes de palabras de la (Figura 29) muestran tres fases diferentes dentro de esta literatura. Al principio, las publicaciones eran más diversas, pero también más generales y orientadas principalmente a los retos técnicos, los impactos y las aplicaciones potenciales. Hoy en día los temas, dominios y fuentes de publicación de esta literatura son muy heterogéneos. Los dominios incluyen la bioinformática, el urbanismo, las ciencias de la computación, la inteligencia artificial (IA), los negocios, la salud y la psicología, entre otros.

En 2015, la investigación se centró en cuestiones técnicas y metodológicas. A finales de la década, el foco de la investigación parece haber pasado de la aplicación de los conocimientos

y técnicas acumulados al desarrollo de técnicas de inteligencia artificial y aprendizaje automático para estudiar problemas específicos en distintos campos.

**Figura 30.** Evolución de las palabras clave en las publicaciones de ciencias sociales sobre big data de 2010, 2015 y 2019 en todo el mundo



*Nota.* Para mejor visualización de los términos, hacer zoom en el documento. Visualización recomendada 150%.

Una revisión de los temas de las publicaciones dentro del campo de las ciencias sociales muestra que a pesar de la importancia que ha adquirido el big data en los últimos años, existe un déficit de producción académica en el papel sobre los efectos que tiene en los procesos de toma de decisiones del estado, el diseño de instrumentos de políticas públicas y su relación con otros sectores de la sociedad (Gamage, 2016; Fredriksson, 2018). La Figura 30 muestra las principales palabras clave de los artículos publicados cada año. El tamaño de las palabras refleja la frecuencia relativa de cada palabra clave dentro de cada año.

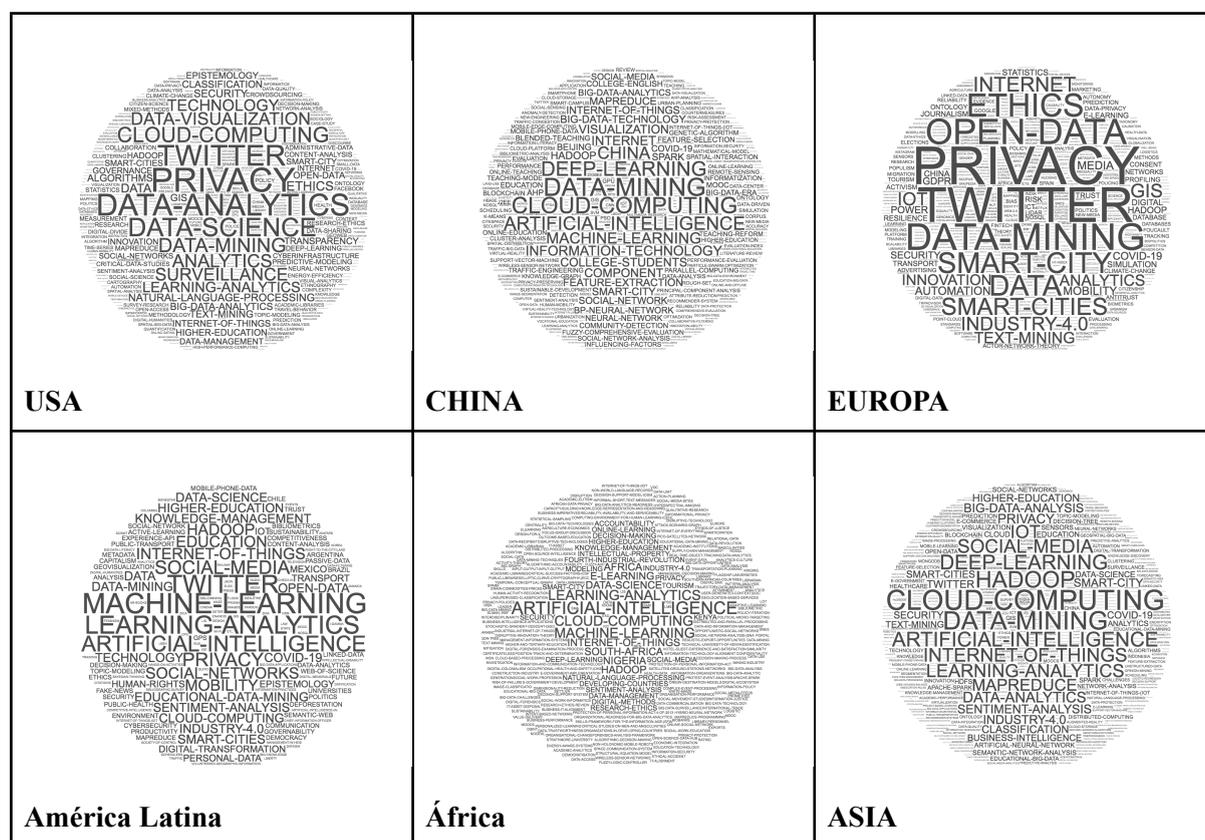
Como vemos, al principio, los pocos artículos sobre ciencias sociales que utilizaban el concepto estaban relacionados con problemas técnicos. Sin embargo, en 2015, gran parte de estas publicaciones parecen haberse centrado en problemas urbanos como el tráfico o las ciudades inteligentes, junto con cuestiones técnicas y metodológicas. En 2019, los medios digitales, la inteligencia artificial, las ciudades y los sistemas inteligentes parecen haberse convertido en las principales preocupaciones de los científicos sociales que utilizan este concepto (Figura 30).

El panorama de los temas principales también cambia si observamos las diferencias entre los países de ingresos agrupados. Las publicaciones de los países de renta alta se centran más en la aplicación de estos conocimientos al aprendizaje automático, la analítica, la informática, la gestión y, en menor medida, a los problemas sociales y de sistemas inteligentes. Los países de

renta media-alta siguen una tendencia similar, pero parecen estar más avanzados en la aplicación de estos conocimientos al desarrollo de sistemas y técnicas de inteligencia artificial.

En cambio, los países de renta media-baja y baja presentan un panorama bastante diferente. Estos últimos se centran en la investigación de cuestiones técnicas relacionadas con los macrodatos, como la minería, el Internet de las cosas, los sistemas inteligentes, etc. Aunque son menos prolíficos, estos últimos países son más diversos y se centran en aprovechar estos conocimientos en el aprendizaje, la predicción, la detección y, en cierta medida, en los problemas sociales (Figura 31).

**Figura 31.** Palabras clave en las publicaciones de ciencias sociales sobre big data por región



*Nota.* Para mejor visualización de los términos, hacer zoom en el documento. Visualización recomendada 150%.

#### ***4.3.7 Palabras claves en los textos Latinoamericanos en las ciencias sociales***

En el contexto específico de América Latina, y considerando los términos y conceptos más utilizados en relación con el big data en las ciencias sociales, se observan algunas tendencias interesantes. Brasil, como uno de los principales exponentes de la investigación en big data en esta región, muestra una diversidad significativa en la elección de palabras y términos, lo que puede deberse a la amplitud de su investigación en este campo. A pesar de esta diversidad, se destacan ciertos términos y conceptos claves como “Cloud-Computing”<sup>46</sup>, lo que sugiere una fuerte preocupación por la capacidad de almacenamiento de información y su acceso a través de la nube.

Luego destaca: “derechos humanos”, “datos personales”, “democracia”, “fake news” y “privacidad”; Estos términos son expresivos y coexisten dentro de una misma temática, que se centra en la preocupación por los posibles impactos negativos que la era de la información, la digitalización y el procesamiento de grandes datos pueden tener en cuestiones relacionadas con los derechos humanos, la democracia, la privacidad y la propagación de noticias falsas.

Al analizar la literatura previa sobre big data y la base de datos Scielo, se ha identificado una tendencia en las ciencias sociales a abordar los riesgos asociados con la implementación de tecnologías de big data en sus disciplinas, y cómo esto puede afectar la vida cotidiana de la sociedad. Además, las entrevistas realizadas con científicos sociales han revelado un sentido de responsabilidad moral entre ellos. Estos académicos consideran que es importante transmitir la idea de que si bien las tecnologías son poderosas y beneficiosas, también pueden introducir sesgos y prejuicios en la sociedad si no se analizan y utilizan adecuadamente:

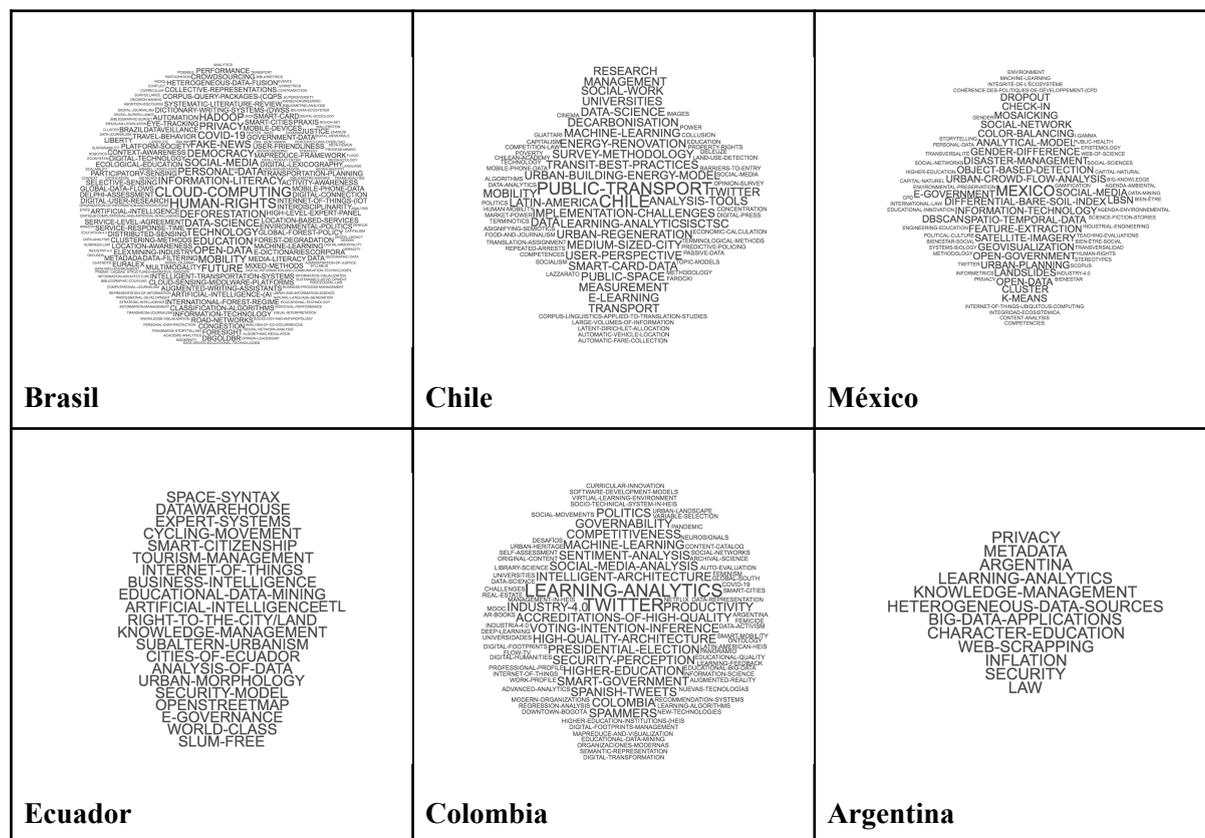
“Llevar la palabra de que las tecnologías son buenas, pero si analizadas de cierta manera, acarrear los mismos sesgos y prejuicios para la sociedad, como lo fue con diversos conocimientos científicos empleados para mal usos, como muchas veces pasó en la historia” (MTP - Entrevista 10/2021 - Brasil - Mi traducción).

Esto refleja la preocupación por evitar que se repitan situaciones históricas en las que el mal uso del conocimiento científico ha tenido consecuencias negativas en la sociedad. Este tema será explorado con mayor profundidad en el capítulo dedicado al ecosistema de ciencia de datos y sus especialistas.

---

<sup>46</sup>Ver glosario.

Figura 32. Palabras clave en las publicaciones de ciencias sociales sobre big data por país



Nota. Para mejor visualización de los términos, hacer zoom en el documento. Visualización recomendada 150%.

Las palabras "education", "open data", "mobility" y "future" forman un grupo de términos que se entrelazan y comparten una fuerza significativa en las discusiones sobre big data en las ciencias sociales. A menudo se encuentran juntos en las publicaciones, lo que sugiere que en el contexto de la investigación están relacionados. Además, se destaca la presencia de un tipo específico de literatura en Scielo: las publicaciones teóricas que especulan y analizan el futuro de las tecnologías y su impacto en la sociedad y en diversas disciplinas.

Un hallazgo importante, respaldado también por el análisis de Scopus, es que "educación" sigue siendo una disciplina activa y altamente interesada en los procesos y la aplicación del procesamiento de grandes datos. Brasil, en particular, se destaca como uno de los países con una producción significativa de textos sobre educación y tecnología, lo que refleja la importancia de esta área en el país.

Asimismo, se observa que en Brasil existe un movimiento considerable en torno al "open data", lo que es relevante dado el contexto de desafíos en el almacenamiento de datos y la necesidad de una solución eficaz. Esta tendencia puede estar relacionada con los términos previamente mencionados, como "Cloud Computing", y con la dimensión y la diversidad poblacional del país. La preocupación por áreas como la movilidad urbana y el tráfico también es evidente, como se ilustra a partir del ejemplo de Río de Janeiro antes mencionado.

En resumen, estas observaciones indican que en Brasil, y posiblemente en toda América Latina, existe un interés sostenido en la intersección entre la educación, la tecnología, el futuro de las tecnologías, la movilidad urbana y el acceso abierto a los datos. Estos temas reflejan las preocupaciones y prioridades de la comunidad académica y pueden tener implicaciones importantes para el desarrollo y la adopción de tecnologías de big data en la región.

En una línea similar de interés pero con una proporción aún mayor, se encuentra Chile, especialmente centrado en las temáticas de tráfico y transporte público. Es relevante observar que en la figura 32, la palabra más destacada es, precisamente, "public transport". A pesar de que Chile, en comparación con Brasil, tiene una población significativamente menor, su sistema de transporte público es ampliamente reconocido en América Latina. La infraestructura de metro, la red de carreteras y las ciclovías hacen de Chile un candidato ideal para la utilización de big data, su interpretación y procesamiento en la gestión urbana y el bienestar público.

Además, se nota una preocupación general por el medio ambiente y temas relacionados con la preservación ambiental. Términos como "decarbonization", "energy renovation", "urban-building-energy model" y "public space" reflejan este interés por abordar cuestiones medioambientales y avanzar hacia soluciones sostenibles, como temas prominentes en la investigación y la discusión en este país.

En contraste, México, Ecuador y Argentina no exhiben un patrón de palabras o términos frecuentes que sea tan claro y definido como en los otros países. Al contrario, estos presentan una diversidad notable en los términos utilizados en sus investigaciones, con la excepción de México, que muestra una leve tendencia hacia aquellos asociados al geoprocesamiento de imágenes satelitales, como "satellite imagery", "geovisualization" y "landslides".

Finalmente, Colombia, como se puede notar presenta mayor interés en términos como “learning analytics”<sup>47</sup> y “twitter”<sup>48</sup>. La relación entre Learning Analytics (analítica del aprendizaje) y Twitter puede ser indirecta, pero aún así, puede haber algunas conexiones importantes.

La primera consistiría en la recopilación de datos educativos en Twitter, lo que lo hacen poder utilizar esta plataforma como una fuente de datos para comprender cómo los estudiantes interactúan con la información relacionada con el aprendizaje. Esto puede incluir el seguimiento de *hashtags* específicos relacionados con cursos, conferencias o temas educativos. Se puede utilizar como parte de la analítica del aprendizaje para obtener una visión más completa de cómo los estudiantes participan en conversaciones educativas en línea (Argonza, 2016).

El segundo sería, por ejemplo, en la comunicación y apoyo estudiantil: Las instituciones educativas pueden utilizar Twitter como una herramienta de comunicación y apoyo para los estudiantes. A través de cuentas institucionales o de docentes, pueden proporcionar actualizaciones, responder a preguntas y brindar información importante (Oguine et al., 2022).

Con relación al mismo tema de educación, Colombia también evidencia términos como “higher-education”, “accreditation-of-high-quality”; y por otro lado una asociación con política: “politics”; “governability”, “presidential-elections”, “smart-government”, demostrando un interés preponderante sobre cuestiones o temáticas políticas relacionada a los datos.

#### **4.3.8 Redes de colaboración**

La figura 12 muestra la evolución de las redes de colaboración internacional durante la última década. La producción científica sobre big data comenzó en varios clústeres centrados en cada país, (EE.UU., China y algunos países europeos) con poca relación entre ellos. Esta característica cambió rápidamente y en 2014 se establecieron diferentes redes de colaboración, con EE.UU. y China como los nodos principales y los países europeos como los principales intermediarios. Esta estructura básica se ha densificado durante los años siguientes, aumentando así las conexiones y llegando a nuevos países y comunidades

---

<sup>47</sup> Ver definición en glosario

<sup>48</sup> Aunque el nombre de la red social actualmente se llama X, en este capítulo, permaneceré utilizando el antiguo nombre de la red social, para preservar el contexto de la investigación respaldada.

científicas. Un dato interesante sobre la estructura y evolución de esta red mundial de colaboraciones es que las instituciones europeas desempeñan un papel fundamental en el desarrollo y circulación de este conocimiento hacia los países de renta media y media-baja.

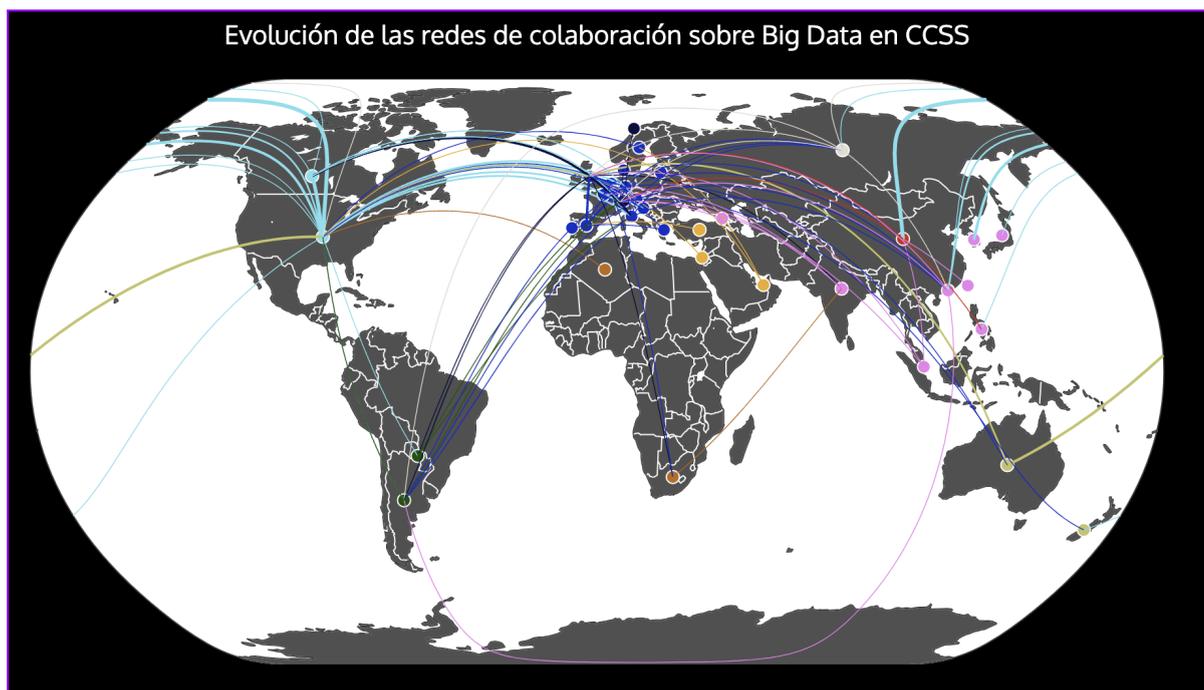
En América Latina, el 48% de los 1.768 vínculos externos registrados entre 2010 y 2019 se establecieron con investigadores afiliados a instituciones europeas, mientras que los norteamericanos representaron sólo el 15%. También fue el caso de África, donde aproximadamente el 40% de las 1.300 colaboraciones externas fueron con países europeos. Las redes asiáticas parecen estar más centradas en sí mismas, con más del 36% de los 16.000 vínculos dentro del continente, pero también tienen fuertes conexiones con Norteamérica (27%) y Europa (25%).

En el siguiente enlace,

<http://viz.divergencelab.org/bigdataSoc/networkEvol/evolRedBDSoc.html#>

[Evolución de las redes de coautoría 2010-2019](#), se puede ver una presentación digital de la circulación del conocimiento sobre big data alrededor del mundo, desde 2010 a 2021. Al clicar en los años correspondientes, en la parte superior del gráfico, se observan los flujos de colaboración en las diferentes regiones del globo. Cada color representa un país. Observé el crecimiento exponencial de colaboraciones a partir del años 2014.

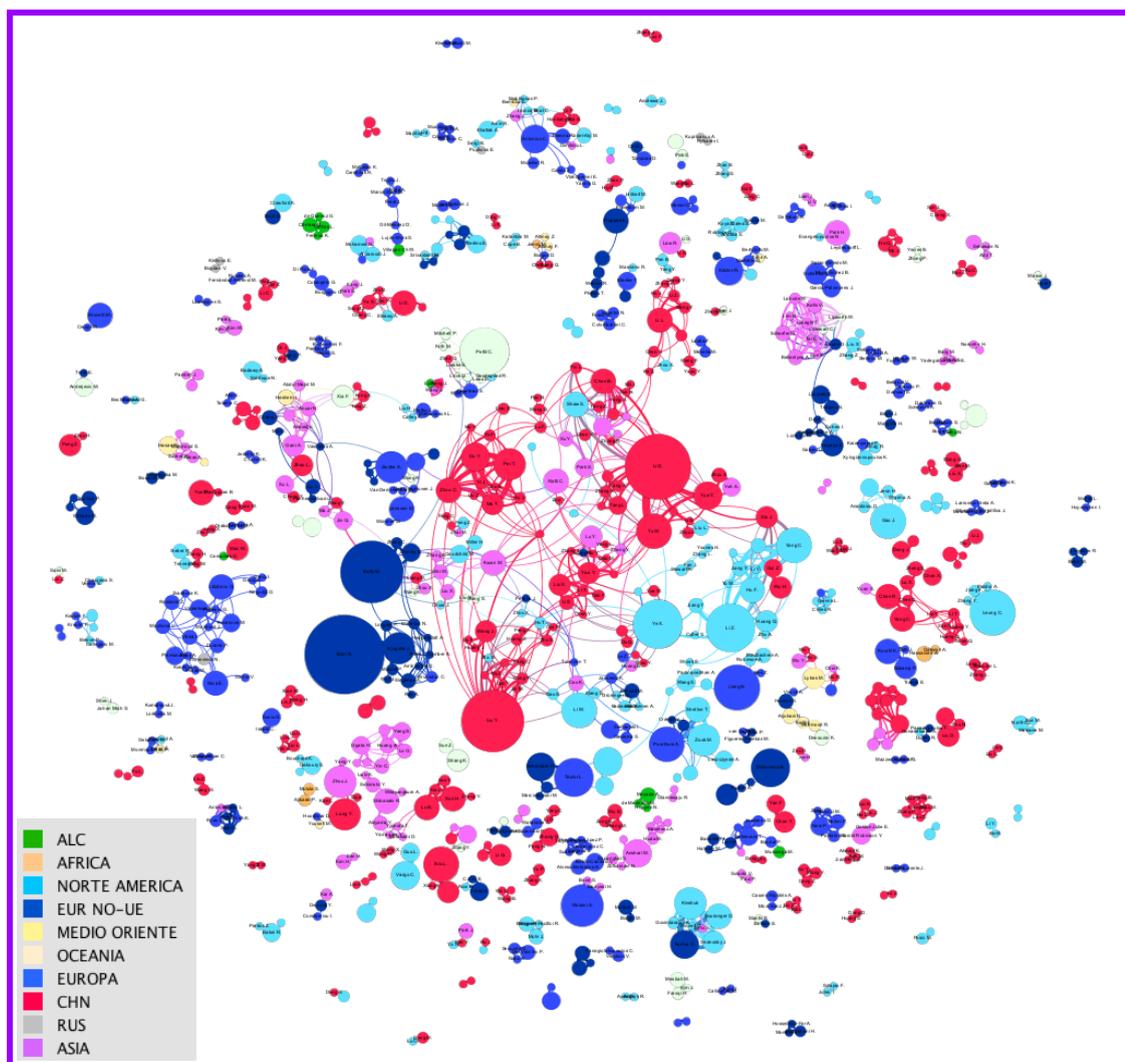
**Figura 33.** Red de colaboración big data en las ciencias sociales en 2014.



Si analizamos las colaboraciones internacionales por autor, obtenemos un panorama relativamente disperso en el que unos 168.000 individuos colaboran mayoritariamente en pequeñas redes independientes: El 62% sólo tiene un vínculo, el 13% tiene más de cinco vínculos y sólo el 5% tiene más de 10 vínculos.

La Figura 33 muestra la red central de la comunidad científica que trabaja en big data, que conecta a unos 1.000 investigadores con más de 10 publicaciones con aproximadamente 9.000 colaboradores en sus países y en el extranjero: El 60% colabora dentro del mismo país y sólo el 18% colabora con autores de países con diferentes categorías de ingresos. La mayoría de estas colaboraciones son entre investigadores chinos, estadounidenses y europeos. También hay colaboraciones entre investigadores sudamericanos e investigadores norteamericanos.

**Figura 34.** *Redes de colaboración internacional por autores*

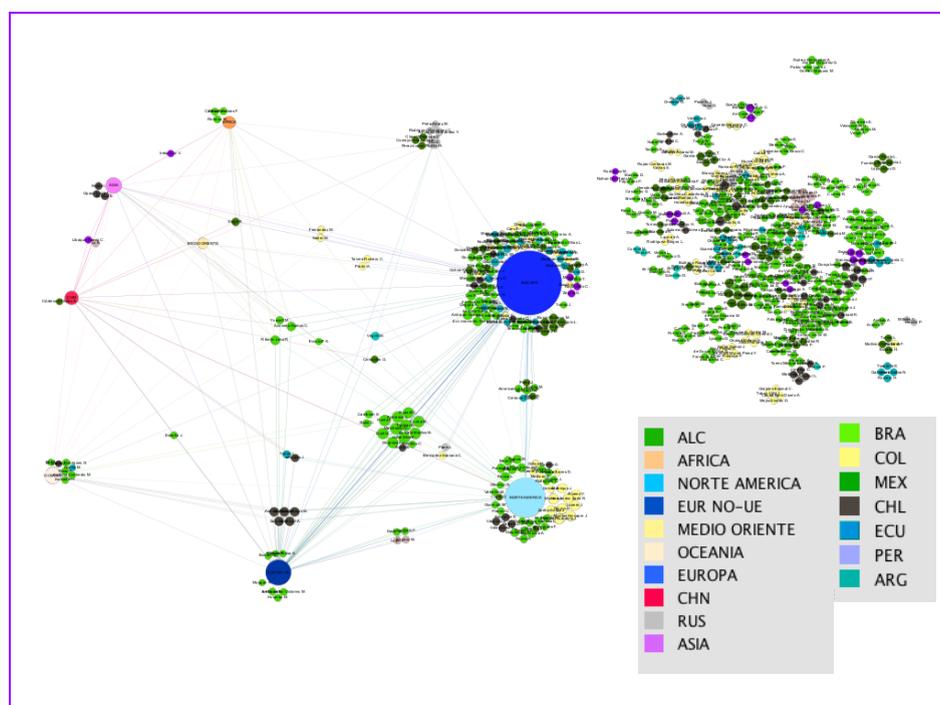


*Nota.* Para visualizar la versión interactiva, acceder a <http://viz.divergencelab.org/bigdataSoc/netBDSocAutMas2Pub.html>

Por último, si observamos las redes de colaboración por institución de afiliación, obtenemos una red global más estructurada, con unos 44.000 actores organizados en torno a dos *hubs*: uno formado por instituciones chinas (17%) y otro por instituciones norteamericanas (18%) y europeas (30%), especialmente universidades (42%). Estos dos polos interactúan directamente a través de una gran variedad de actores más pequeños, incluidas instituciones de distintos continentes, regiones y categorías geoeconómicas. La institución media tiene aproximadamente 100 vínculos con otras instituciones y las más grandes (<1%) tienen más de 1.000 vínculos.

La figura 35 muestra la red de instituciones de afiliación que colaboran en la investigación sobre big data con instituciones del mismo país (47%) y del extranjero (53%). Entre estas últimas, sólo el 17% colabora con instituciones de diferentes categorizaciones según ingresos. Las instituciones europeas tienen vínculos con instituciones africanas, latinoamericanas y asiáticas. Los datos también muestran algunas colaboraciones sur-sur entre países de América Latina, África y Asia.

**Figura 35.** *Redes de colaboración internacional por instituciones*



*Nota.* Para visualizar la versión interactiva, acceder a <http://viz.divergencelab.org/bigdataSoc/netBDSocAutReg.html>

#### ***4.3.9 Conclusiones observadas a partir del estudio bibliométrico en Scopus***

Hemos presentado una cartografía analítica de la investigación sobre big data a partir de un conjunto de más de 73.000 entradas de la base de datos Scopus publicadas en las dos últimas décadas. Evaluamos este corpus en tres aspectos principales: (1) productividad científica e indicadores de rendimiento; (2) principales áreas de investigación y tendencias temáticas y (3) redes de colaboración entre autores, instituciones de investigación y países de afiliación; prestando especial atención a las principales relaciones y diferencias entre las comunidades científicas que trabajan en este tema en países situados en regiones y condiciones económicas diferentes.

Nuestras principales conclusiones muestran que la productividad científica ha aumentado exponencialmente desde 2010, con una tasa media de crecimiento de aproximadamente el 90% anual. China y Estados Unidos lideran la producción científica sobre big data. Según el nivel de renta del país, los chinos, sudafricanos y colombianos lideran en el grupo de los países de renta media-alta, mientras que India y Marruecos lideran entre los países de renta media-baja.

De las 20 principales instituciones que acogen a investigadores en big data, 18 son chinas, seguidas de la Universidad de Delhi y el MIT. Lectures Notes in Computer Science es el ágora más importante sobre la temática, e IEEE Access es la principal revista académica que trata sobre los datos masivos.

En cuanto a las citas, además de la mayor productividad de los investigadores e instituciones chinos en este campo, las contribuciones estadounidenses siguen siendo las más influyentes y los artículos y autores más citados proceden de EE.UU. A pesar del dinamismo del campo, alrededor del 2% de los artículos concentran el 40% de las citas del campo, mientras que el 42% de estas publicaciones no tienen ninguna cita.

Los principales campos científicos que publican sobre big data son informática e ingeniería (81% de las publicaciones en todo el periodo), seguido de medicina y ciencias sociales. Sin embargo, desde 2017, la importancia relativa de esta área se ha reducido al 35% debido al desarrollo de proyectos interdisciplinarios en matemáticas, ciencias sociales y ciencia de la decisión, medicina y negocios.

La tendencia de las palabras clave a lo largo del tiempo muestra que, en 2010, la literatura estaba principalmente orientada a los desafíos técnicos, los impactos y las posibles

implementaciones de estas tecnologías. En 2015 se centró en cuestiones técnicas y metodológicas, mientras que en los últimos años se ha desplazado hacia el desarrollo de técnicas de IA y aprendizaje automático.

Por último, observamos una importante actividad de colaboración científica: sólo el 12% de las publicaciones son contribuciones individuales. Sin embargo, la mayoría de las redes de colaboración son nacionales; el 19% pertenece a redes internacionales. En los últimos cinco años se ha pasado de agrupaciones centradas en los países a una red más internacional, en la que Estados Unidos y China son los principales nodos. Los países europeos parecen ser los principales intermediarios en la circulación y el desarrollo del conocimiento en este campo con países de África y Sudamérica. Aunque en menor medida, también hemos detectado algunas colaboraciones sur-sur entre América Latina, África y Asia.

Así, he presentado una caracterización detallada y un análisis exhaustivo de la investigación global sobre big data en la última década. Esta investigación actualiza trabajos bibliométricos anteriores ampliando el corpus analizado previamente y explorando una base de datos no investigada.

El aporte más importante al análisis bibliométrico son las ideas que se entregan sobre las diferencias en la productividad científica, las áreas de investigación y los temas de tendencia, así como las redes de colaboración entre países de distintas condiciones geoeconómicas. Estas diferencias ponen de manifiesto la desigual distribución y circulación del conocimiento sobre big data que subyace al crecimiento de las publicaciones en la última década. Nuevas investigaciones podrían proporcionar una caracterización más profunda y detallada de la investigación en este campo en regiones y países concretos, así como en áreas y temas de investigación específicos.

## Cap 5: Scielo y Scopus: que los textos nos tiene a decir

Este capítulo surge como una extensión de la investigación bibliométrica inicial, en respuesta a la inquietud de profundizar en los contenidos de los textos analizados. Dado que la base de datos, sobre todo en Scielo, es pequeña, resultó viable llevar a cabo un análisis más detallado de cada publicación, resultando en este capítulo totalmente dedicado a la interpretación del contenido de los textos estudiados. Aquí, los textos están organizados por disciplina, lo que facilita al lector comprender las distintas perspectivas y enfoques presentes en las publicaciones seleccionadas. Este capítulo, por lo tanto, complementa el análisis bibliométrico inicial al integrar una dimensión cualitativa, ofreciendo una visión más completa que va más allá de los datos cuantitativos, captando y demostrando así tendencia de temáticas y contenidos en el campo de estudio investigado.

El contenido aquí presente, es una especie de resumen del estado del arte de toda mi búsqueda bibliográfica. Fueron estos textos los que me guiaron y orientaron la entrada al campo en la fase etnográfica de la investigación, posibilitando así, generar al estudio nuevas categorías de análisis, presentada más específicamente en los capítulos 8 y 9 de este trabajo.

De manera general, a título de proporcionar una introducción más abarcativa a cerca de los textos, queda claro que a comienzos de los años 90, el big data se aplicaba casi exclusivamente al contexto de los mercados digitales, agencias de publicidades y redes de consumo (Gómez-Barroso & Feijóo-González, 2013). La popularización del término estaba, entonces, íntimamente asociada a la difusión masiva de los computadores y sistemas computacionales, y a la emergencia de lo que muchos autores consideran como la cuarta revolución industrial (Schwab, 2016) o la era digital (Castells, 1996). Su popularización es consecuencia de la magnitud de información y conocimiento variados de casi toda actividad humana y no humana que pasamos a crear en los últimos 30, 20 años: el poder computacional, de memoria, redes, comunicación y sensores, es en definitiva el catalizador de este proceso.

La lectura de estos textos, evidenciaron que a partir de los años 2000, más específicamente 2004, se observa el inicio de un nuevo marco de difusión del término, esta vez asociado al campo científico (Monleon-Getino, A, 2015)<sup>49</sup>: la era digital funcionó como catalizador de la

---

<sup>49</sup> Esta información puede corroborarse a través de Google Trends donde en 2004 se percibe el primer pico de búsqueda de “big data” + “science”:  
<https://trends.google.com/trends/explore?date=all&q=big%20data%20science>, accedido en 2 de julio de 2021.

producción de datos, lo que contribuyó al establecimiento del valor prominente de los datos. La información asume, entonces, el centro de la economía global (OCDE, 1996).

Es a partir de esta gran disponibilidad de registro y acceso a los flujos de datos que *big data*, que se vió cada vez más consolidado como una tendencia en el mundo científico. La academia, desde entonces, fue apropiándose del uso duro de los datos para entender, investigar y razonar sobre el comportamiento de materiales tanto biológicos como físicos, de materias primas, y por supuesto, para comprender a los distintos grupos sociales y la sociedad como un todo (Rocha, 2018).

Teniendo en cuenta tal realidad, a partir del 2014 observase el surgimiento de la primera revista académica dedicada específicamente a la temática “Big data y Sociedad” (*Big Data & Society, BD&S, SAGE*)<sup>50</sup>. También se percibe a partir de las herramientas de búsquedas digitales un crecimiento de más de 100% por los términos “big data and social science”<sup>51</sup>.

El interés por relacionar big data y el estudio de las sociedades puede estar vinculado a diversos factores. El primero de ellos está asociado a la propia naturaleza de este tipo de datos: los macrodatos, que no son solamente un banco de datos gigantesco. Antes que nada son datos digitales. Eso significa que en la propia constitución de una información está suscrita la síntesis de los datos a partir de la codificación 0 y 1, lenguaje propio de los computadores. A diferencia del lenguaje humano, es más simplificado, permitiendo que el proceso de producción y captura de datos se haya potencializado tanto en cantidad como en variedad de contenido.

Así, el crecimiento exponencial de la capacidad de producir y capturar datos (digitales) a través de la accesibilidad cotidiana de aparatos y sistemas de registro constantemente conectados a las personas —relojes, tablets, notebooks, celulares, cámaras y sensores— hace que sea posible el almacenamiento de miles de millones de datos y que estos sean manipulados, o procesados, de manera rápida y universal, ya que el nivel de traducción entre

---

<sup>50</sup> Periódico digital: <https://journals.sagepub.com/home/bds>.

<sup>51</sup> Dato proveniente de Google Trends, donde las búsquedas del periodo anterior a 2013 suman 25 puntos. A partir de 2013, 64 puntos <https://trends.google.com/trends/explore?date=all&q=big%20data%20social%20science>, accedido en 2 de julio de 2021.

los variados datos digitales se vuelve más sencillo debido a las características del lenguaje computacional (Gutiérrez, 2018).

El almacenamiento de este flujo constante e ilimitado de información establece una presencia real, y muchas veces definitiva, en el espacio digital de la vida de las personas. Tal escenario propicia, por consecuencia, un ambiente fértil para la investigación social vinculada a los macrodatos: la digitalización de la información, viene contribuyendo con un aumento significativo de la capacidad de generar registros y materialidad alrededor de la vida social, potencializando así la generación de datos sobre las costumbres y prácticas sociales (Bastin & Tubaro, 2018).

La viabilidad de transponer al espacio digital cualquier tipo de registro producido en lo que comúnmente llamamos de “mundo real”, instaura una capacidad de transitar entre estas dos “realidades” de forma muy semejante, o incluso con más detalles, que en el “mundo real”. En otras palabras, para investigadores como Daniel Miller (2012) el mundo virtual es una copia digital del mundo material donde vivimos. Es por eso que la información generada en contexto de la digitalidad debe ser comprendida como dato verosímil sobre la realidad tanto de los individuos, como de grupos sociales o, incluso, de toda la sociedad. La tendencia creciente en entender las sociedades y los grupos sociales a través de este vasto universo informacional que se nos viene presentando genera inquietud, una curiosidad natural y esperada al interior de las Ciencias Sociales. La incorporación del procesamiento de grandes cantidades de datos como forma de obtener información sobre la sociedad, abre todo un nuevo campo de aplicación y tratamiento de las informaciones sobre los sujetos y los grupos sociales en el contexto investigativo social (Nascimento, 2016).

De manera general, en los textos se percibe la difusión de la idea de que big data se presenta como una innovadora técnica capaz de optimizar procesos de investigación, acortando los tiempos y dificultades de procesar grandes cantidades de datos. Nos resta saber e investigar, que tan promisorio, o hasta incluso “revolucionario” podría ser las tecnologías de grandes procesamientos de datos y cómo afectan a la producción de conocimientos sociales latinoamericano y su comprensión alrededor de las representaciones sobre el social, objetivo mayor de esta investigación.

En la secuencia está presentado un resumen de los contenidos encontrados a partir de la lectura de las publicaciones analizadas como resultado del estudio bibliométrico hecho.

Iniciaremos por abordar la temática del surgimiento del procesamiento computacional de texto como una tecnología que viabiliza la manipulación y existencia de los macro datos, a partir, es claro, de la visión de los estudios de ciencias sociales, entrando por consiguiente, a la presentación de los contenidos trabajados en estos textos, disciplina por disciplina.

### ***5.1 El procesamiento computacional de textos como el origen de la manipulación de macrodatos sociales***

Dos de los primeros textos con los que tuve contacto fueron “Adapting Computational Text Analysis to Social Science (and vice versa)” de DiMaggio (2015) y “Big Data and Historical Social Science” de Bearman (2015), ambos publicados por la revista *Big Data & Society*. Estos dos autores fueron los primeros que aparecieron en mi lista y creo que no por causalidad. Aunque historia como disciplina no pertenece oficialmente a la lista que componen las ciencias sociales, de acuerdo con OCDE (referencia utilizada para delimitar el campo de la ciencias sociales en este texto) no pude dejar de percibir la marcada presencia de los estudios históricos, o de historia, involucrados en el tema de big data.

Conforme investigaba, me fue quedando claro la importancia de la utilización de los macrodatos para los estudios en historia, y esto se evidencia en la frecuencia y relativa facilidad para encontrar artículos y textos académicos alrededor de esta disciplina<sup>52</sup>. Sobre todo por una dimensión bien peculiar: la posibilidad de contar con herramientas computacionales de interpretación, estudio y análisis de textos y documentos, evidencias tradicionalmente muy caras en los estudios de Historia, de una forma general.

El procesamiento computacional de textos, sin duda es un gran recurso disponible para las ciencias sociales. Se está usando en diversas disciplinas de este campo científico, para los más variados fines y aplicaciones. Por esto también, según la percepción como especialista que he tomado en estos temas, es que es más común que los textos encontrados en este primer acercamiento discurren sobre, o están aplicando en sus investigaciones científicas en historia<sup>53</sup>.

---

<sup>52</sup> En mi caso, 4 de los 13 textos de “ciencias sociales” + “big data”, representan el 30% de los textos que variaron entre 10 disciplinas distintas.

<sup>53</sup> Lo que pude comprobar en mi participación en el III Congreso Internacional HD Rio Humanidades Digitais, 2023. Transcribo una nota de mi cuaderno de campo, del día 19 de Abril de 2023: “Hoy salimos para tomar cerveza con una galera do evento. Eramos 7 en una mesa: 3 de historia, 2 da TI, 1 de las ciencias políticas y yo de la “sociología”. Enquanto conversamos, percebi que o congresso tenia 3 dias completos, com cronograma de manhã à tarde, como mesas temáticas provenientes de diferentes ramos da história. Enquanto meu eixo, 8

El artículo de DiMaggio (2015), es un ejemplo de ello y se centra en las potencialidades del procesamiento computacional del análisis de textos, práctica que se ha popularizado cada vez más entre los científicos sociales, según el mismo autor. Él plantea, entonces, la necesidad de que investigadores sociales y computacionales unan fuerzas con el objetivo de mejorar las técnicas de análisis textual y lenguaje natural, habitual en las investigaciones sociales. Según el autor, el aprendizaje automático y las técnicas big data pueden contribuir a mejorar la objetividad de las codificaciones siempre y cuando la implementación de los algoritmos por detrás de cada etapa técnica sea trabajado conjuntamente con las interpretaciones sociales. Y alienta fuertemente a que las demás disciplinas de las humanidades incorporen el hábito de procesar sus textos de manera computacional. De la lectura de este artículo me inspiré para probar algunas técnicas de procesamiento de textos durante mi posterior estudio bibliométrico.

En él se evidencia, además, que mientras los científicos sociales suelen obsesionarse con las causalidades y se basan en pruebas formales de significación estadística, los informáticos acaban por priorizar un análisis de texto centrado en los resultados. No abogando en favor de uno u otro, el autor defiende que la colaboración entre científicos sociales y científicos de la computación es esencial para mejorar las técnicas de análisis de textos, pues la combinación de ambos enfoques posibilitaría una mayor variabilidad de modelos, lo que agregaría diversidad y profundidad a los análisis textuales. Por otra parte, al ser capaces de transitar entre diversos tipos de modelos, es probable que logremos establecer mejores reglas generales para ayudar a los analistas de textos a mejorar sus modelos y así, encajar más adecuadamente las especificidades de nuestras investigaciones.

La implicación de la introducción de un saber big data al procesamiento y análisis de texto a través de técnicas computacionales en ciencias sociales es el hecho de que los conjuntos de datos pueden ser repensados por medio de nuevas estructuras de las conexiones causales. Como consecuencia, estaríamos creando las bases para la consolidación de nuevas estrategias, patrones y flujos comportamentales a la hora de analizar textos científicos y académicos (DiMaggio, 2015). Al combinar diferentes perspectivas y habilidades, podemos aprovechar al máximo el potencial de ambos campos y avanzar en nuestra comprensión de los datos textuales en las ciencias sociales, y por ende, de la propia sociedad.

---

expositores no total, meio dia de apresentação no último dia do congresso! Me lembrei da "coincidência" da minha primeira busca bibliográfica. apareciam em maior números os artigos de história quando pesquisava BD+CS. Não é por acaso. Eles amam o texto. E são muito bons em estruturar o pensamento. Fiquei curiosa. Bom pra pensar”.

Sobre el procesamiento de lenguaje natural y las potencialidades de usar big data para el análisis textual, está el texto de Sophie Mützel (2015), quien plantea que contribuyen al contexto de investigación en sociología al proporcionar una gran cantidad de datos. Estos datos podrían, entonces, ser analizados con técnicas avanzadas para obtener información valiosa sobre la sociedad. Por ejemplo, los investigadores pueden utilizar técnicas de minería de datos<sup>54</sup> y análisis de texto para identificar patrones y tendencias en grandes conjuntos de datos, lo que puede ayudar a comprender mejor cómo funciona la sociedad. Además, estos datos también pueden ayudar a los investigadores a realizar estudios longitudinales y comparativos a gran escala, lo que puede proporcionar una visión más macrosocial y precisa de los fenómenos sociales. En resumen, el uso de big data puede contribuir a la sociología con nuevas formas de recopilación y análisis de datos, lo que mejoraría nuestra comprensión del mundo social.

Por otra parte, el autor explica cómo la investigación a partir de técnicas big data suele ser un buen inicio para sondear categorías de análisis de texto, ya que la utilización de los *topic modelings* favorece la visualización de temas, asuntos y términos más frecuentes y de mayor peso, para el banco de datos textual trabajado. Entre los beneficios de los métodos de procesamiento computacional de textos, los “modelados de temas”, está a) la posibilidad de analizar procesos emergentes desde una visión macroscópica utilizando grandes corpus de texto; b) la búsqueda por temas latentes sin necesariamente tener una codificación de los materiales a priori, lo que contribuye a una mayor imparcialidad en cuanto al análisis de contenido de los textos; c) la posibilidad de seguir las trayectorias discursivas, pudiendo así, describir y rastrear desarrollos textuales durante largos periodos de tiempo (Mützel, 2015).

### ***5.2 La historia: reinterpretando el pasado a través de técnicas big data***

Entre los textos sobre historia, destacó el artículo de Bearman (2015), donde discute cómo implementar big data a las investigaciones históricas. Entre ellas, se podría cambiar la manera en que accedemos a los documentos, textos y registros de distintas dimensiones de nuestro pasado. La posibilidad de visitar archivos (por medio de la digitalización de ellos), logra que, junto a los datos ya existentes, se hagan más accesibles en menos tiempo y, lo más importante según Bearman, puedan cruzarse con otros tipos de datos (geolocalizados, sensores, fotografías, notas, etc) a fin de abrir nuevas interrogantes sobre cómo hemos

---

<sup>54</sup> Ver el glosario, al final de este texto.

concebido los relatos históricos, hasta ahora, dados como hechos consolidados (Bearman, 2015).

El uso de big data permitiría, entonces, reimaginar el pasado, reorganizando, así, el modo en que comprendemos el presente. Podría, también, ser una nueva manera de recolectar datos, lo que generaría nuevos u otros tipos de datos, generando así, nuevos relatos, nuevas narrativas, categorías e incluso, nuevas teorías históricas. Así, se puede repensar la modernidad con la nueva información sobre hechos históricos que surgen de la recolección y recombinación de estos nuevos/otros tipos de datos como es el caso la comprensión de la esclavitud americana por medio de los archivos del comercio negrero de una ciudad, subjugado a datos georeferenciales (Bearman, 2015).

Para las ciencias sociales en general, esto significaría la posibilidad de indexar relaciones sociales en los registros de datos, lo que entregaría la potencialidad de reorganizar la historia social a través de los llamados “*transposable mechanisms*”: mecanismos orientados para responder preguntas donde la causalidad es lo que importa, no así las narrativas, como se acostumbra a hacer en Historia. El foco estaría en identificar el ritmo de las estructuras de cambio que surgen en los entornos institucionales claves, conocidos por sustentar grandes procesos históricos de interés.

### ***5.3 Economía y negocios: una interpretación macro social de la vida económica a través de big data***

Otra disciplina responsable por la popularización de técnicas de big data es la economía, particularmente en el área de los negocios. En cualquier búsqueda rápida sobre el tema en internet encontramos indicios de que usar macrodatos en los estudios de economía ha sido una tendencia creciente en los últimos años.

Según el texto más citado de acuerdo a Google Scholar relacionado a big data y economía de Binbin Wang, Xiaoyan Li, (2017) el procesamiento computacional de grandes cantidades de datos es utilizado en economía para rastrear el estado de las transacciones del mercado, instruir y actualizar en tiempo real los índices de oferta y demanda de los principales productos y servicios. Además, proporcionan una referencia para la toma de decisiones de producción en la industria. Son también utilizados para planificar el tipo y la cantidad de bienes y servicios ofrecidos, así como la combinación de factores técnicos y el

posicionamiento en el mercado de estos, basado en los catálogos flexibles de orientación de producción de determinados bienes.

Otro tipo de aplicación apuntado por las lecturas es la utilización del análisis estadístico de big data para gestionar datos personales y tomar decisiones en la gestión económica a través de la recopilación de datos de diversas fuentes, como redes sociales, transacciones en línea, sensores, registros gubernamentales y más. Estos conjuntos de datos grandes y complejos proporcionan una visión más amplia y detallada de la actividad económica (LuQuan Wang, 2020).

La literatura científica económica centrada en el campo empresarial, es otro espacio importante de aprovechamiento de big data. Las tecnologías de procesamiento computacional de grandes datos puede apoyar y proporcionar logística de datos capaces de transformar los procesos empresariales corrientes en servicios financieros bastantes complejos como en la detección de fraudes, el análisis de riesgos y la negociación algorítmica en tiempo real (Basdas, Esen, M.F. 2021).

En el universo de la Economía con E mayúscula, la aplicación de métodos de aprendizaje automático para gestionar macrodatos en economía y econometría, es una tendencia cada vez más adoptada por los gestores de las grandes economías nacionales. Mediante la minería de textos<sup>55</sup> y la previsión de pronósticos, el uso de macrodatos permite a los economistas generar pronósticos y predicciones más precisas sobre variables económicas clave, como el crecimiento del PIB, la inflación, el desempleo y las tendencias del mercado (Varian, 2014). Al aprovechar grandes cantidades de datos en tiempo real, los modelos pueden mejorar su capacidad para predecir eventos económicos con mayor precisión, entre tantas otras cosas.

De manera general, a través de la lectura de los artículos seleccionados, pude percibir que la tónica de los textos en economía evidencian una apuesta por mejorar la gestión de políticas económicas y de mayor eficiencia en la toma de decisiones en el ámbito público y privado, ya que los macrodatos posibilitarían un entendimiento más complejo sobre el comportamiento

---

<sup>55</sup> La minería de textos es un campo de estudio interdisciplinario que combina técnicas de procesamiento del lenguaje natural, aprendizaje automático y minería de datos para extraer información valiosa y conocimientos útiles de grandes volúmenes de texto no estructurado. El objetivo principal de la minería de textos es descubrir patrones, tendencias, relaciones y conocimientos implícitos en los documentos de texto, como artículos, libros, correos electrónicos, redes sociales, informes y cualquier otra forma de contenido textual. Al aplicar técnicas de minería de textos, se puede realizar una exploración automática y sistemática de los datos textuales para obtener información relevante y significativa (Arévalo, 2018) Ver en: <https://universoabierto.org/2018/02/22/que-es-la-mineria-de-textos-como-funciona-y-por-que-es-util/>

económico tanto a nivel de los usuarios-individuos, como al de usuarios colectivos y consumidores (Chavan y Akarte, 2014). Esta minuciosidad de análisis, permitiría diseñar intervenciones más efectivas y tomar decisiones más informadas en áreas como la regulación económica, la inversión y el comercio. El análisis y modelados de grandes cantidades de datos económicos ayudaría a los economistas a comprender mejor el comportamiento económico de diversos grupos sociales.

#### ***5.4 La psicología y la intervención en salud a partir de big data: recalibrando la disciplina***

La reflexión sobre la relación entre macrodatos y psicología en los textos encontrados hablan respecto a la influencia de las técnicas de análisis de grandes datos en ámbitos como evaluación psicológica, intervención en salud, investigación y docencia.

Se analiza, por ejemplo, cómo el uso de grandes bases de datos y tecnologías como los smartphones, puede ser una oportunidad para mejorar la práctica clínica? de la psicología y contribuir a un cambio de era en esta disciplina (Ruiz, et al. 2015), al aumentar el nivel de personalización de las intervenciones. Como consecuencia, se observaría un aumento en la participación de los pacientes, lo que posibilitaría una mayor exactitud en la predicción de la conducta. Permitiría, además, una evaluación más precisa y objetiva de los resultados de investigación y mejoraría la docencia en Psicología, para dar otro ejemplo presente en el texto

En un segundo texto por Nombre Becerra (2018), se menciona que gracias a la facilidad y flexibilidad en la recogida de datos, se va a poder estrechar la clásica brecha entre la psicología aplicada y la académica. Sin embargo, señala que para ello se requerirá un trabajo conjunto que permita el avance disciplinar en temas tan concretos como la acreditación y homologación de aplicaciones psicológicas que cuenten con todas las garantías científicas, desde su eficacia, eficiencia y utilidad hasta en su capacidad de contribuir más datos (Becerra, 2018).

Otra utilidad de big data mencionada es la psicometría, el que utiliza macrodatos y redes sociales para medir el comportamiento de los individuos a través de mecanismos digitales. Los métodos y procedimientos de estas mediciones están basados en la correlación entre signos y huellas alojados en bases de datos y valores que se les asignan y que dan cuenta de rasgos medibles (Gómez Barrera, Arango Tobón, 2022).

En los textos recolectados y leídos, es unánime la noción de que los macrodatos evidencian un potencial para proporcionar valor a la psicología. Sin embargo, también se percibe la preocupación, o responsabilidad teórica, en apuntar de manera crítica las limitaciones y cuidados debidos al implementar las técnicas big data en las labores de la disciplina, sean prácticos o científicos (Becerra, 2018). La búsqueda de big data sigue siendo una empresa incierta y arriesgada para el investigador psicológico promedio (Adjerid, 2018). Parte de la incertidumbre en la discusión sobre el potencial impacto de los macrodatos es el desafío que significa para los psicólogos mantener un alto nivel teórico y crítico frente a la fascinación que el impacto de la utilización de tecnologías big data pueden provocar (Ruiz et al., 2015).

### ***5.5 Las ciencias de la educación y los nuevos procesos de enseñanza a partir de big data***

La educación es otra de las disciplinas que viene difundiendo las aplicabilidades de las tecnologías big data en pro de su desarrollo. El texto más “antiguo” encontrado en mi primera búsqueda en Google Scholar es bastante optimista en cuanto a la introducción de la computación en la investigación científica de la disciplina. Hay un cierto ánimo con relación a posibilidades de que la investigación educativa acabe creciendo y mejorando como consecuencia (Argonza, 2016).

Según el autor, el uso de big data en la ciencia de la educación ha permitido el avance de la investigación educativa. Los investigadores, pudiendo acceder a grandes cantidades de datos y analizarlos, pasan a obtener nuevos conocimientos sobre los procesos de enseñanza y aprendizaje, lo que posibilita crear nuevos mecanismos de práctica docente, e inclusive nuevas maneras de evaluar la efectividad de diferentes enfoques educativos (Argonza, 2016).

Con relación a la práctica educacional, en las aulas, recurrir a los macrodatos puede proporcionar información relevante para aportar a la calidad de la educación y el rendimiento de los estudiantes, estudiando las correlaciones entre variables educativas, evaluación de estrategias pedagógicas, resultados académicos y la influencia de diferentes factores externos al ambiente escolar (horas de sueño, alimentación del estudiante, nivel de estrés, entre otros) de forma conjunta (Oguine, Oguine K. y Bisallah, 2022). Este tipo de estudios son llamados de identificación de factores de éxito estudiantil, y ayudan a los educadores a identificar patrones de comportamiento, intervenciones efectivas y prácticas pedagógicas que conduzcan a una mejoría en el desempeño educacional. Se basan en la recopilación y el análisis de datos

a gran escala tales como registros de calificaciones, datos demográficos, interacciones en línea y otras fuentes relevantes (Oguine, Oguine K. y Bisallah, 2022).

Otro tipo de implementación de big data en educación apuntada en los textos incluye la detección temprana de estudiantes en riesgo de abandono (Estrada et al, 2020). Al analizar grandes conjuntos de datos, se pueden identificar señales tempranas de problemas académicos o dificultades de aprendizaje, permitiendo a los educadores intervenir de manera más asertiva en base a cada caso y proporcionar apoyo adicional a los estudiantes que lo necesitan.

En términos generales, el énfasis en traer a la luz la personalización del aprendizaje como una potencialidad proveniente de la introducción de big data al escenario de la investigación científica, fue la temática más discutida en los textos encontrados. Como consecuencia, se presenta una derivación un tanto seductora: la mejora de la eficiencia operativa de las instituciones educativas y la toma de decisiones basadas en datos cada vez más asertivos (Didik Madyatmadja et al., 2021). La lógica por detrás de este raciocinio es que el análisis de datos masivos puede ayudar a identificar patrones y tendencias en el comportamiento de los estudiantes, lo que puede ser útil para mejorar la enseñanza y el aprendizaje. Al personalizar la experiencia de aprendizaje de cada estudiante sería posible, entonces, comprender, a partir de más elementos, los estilos de aprendizaje individuales, al mismo tiempo en que se acompaña y observa a estos jóvenes de manera colectiva, al identificar patrones en las preferencias no aisladas.

### ***5.6 La sociología y su discusión crítica cuanto a la introducción de técnicas big data en el desarrollo del saber social***

Como podría intuir, en el caso de los textos de sociología, es bastante evidente la mayor presencia de una postura crítica o cautelosa respecto al impacto de big data en esta disciplina. Los textos en sociología tienden a asumir una postura expectante en cuanto a las implementaciones de las tecnologías big data, y poseen los textos de contenido más teórico-epistemológico con relación a los impactos posibles de la entrada de las formas computacionales al investigar la información sociológica.

Podemos verlo en Boyd y Crawford, (2011) y Maestro Cano (2019) donde se puede percibir la existencia de una cierta tradición, o vocación, por centrarse en las divisiones disciplinares

adyacentes a la disciplina a partir de la introducción de big data a las actividades científicas sociológicas. Son textos que problematizan la emergencia de nuevas metodologías de investigación al quehacer científico y la producción de conocimientos disciplinarios, demostrando cierta apertura y comprensión de las transformaciones provenientes de un nuevo contexto. También son los pioneros en discutir los peligros, los enredos, o las falsas promesas, que, en nuestro caso, big data y la introducción de procesamiento computacional de datos, puede instaurar en los procesos investigativos de las sociedades.

Con todo, es evidente que, conforme se va actualizando la literatura, el reconocimiento de que la utilización de big data en sociología ha abierto nuevas oportunidades para comprender los patrones y tendencias sociales (Tinati, Halford, Carr y Pope, 2014). En el campo de la sociología, el big data se utiliza para analizar y comprender una amplia gama de fenómenos sociales. Estos datos pueden provenir de diversas fuentes como redes sociales, transacciones en línea, registros gubernamentales y sensores, entre otros.

También son textos más reflexivos en cuanto a la problematización de conceptos asociados a big data, como: “datificación del mundo”, “digitalidad”, “modernidad”, “inteligencia artificial” y “algoritmos” (Bastin & Tubaro, 2018). Estas discusiones teóricas casi siempre culminan con la discusión de la implicación de utilización de big data y las posibles consecuencias para el desarrollo de la propia disciplina y el mundo social.

Entre las disciplinas levantadas en esta búsqueda inicial de referencias bibliográficas, la sociología, sin duda, es la que menos presenta textos provenientes de investigación aplicada. Y junto con el derecho, son el área que más textos tienen sobre las implicaciones teóricas, epistemológicas y metodológicas de un futuro científico computacional.

El autor que explora las potencialidades reales de utilización de big data en las ciencias sociales es Martínez-Uribe (2022), quien evidencia cómo las publicaciones en revistas de impacto pueden ser analizadas a través de grandes colecciones de datos para entender mejor las relaciones entre investigadores, instituciones, revistas y editoriales en el campo de la sociología.

Según él, con la posibilidad que ofrece big data, las grandes colecciones de datos permiten la representación de los vínculos sociales que se dan en la ciencia. Las grandes fuentes de datos bibliométricos, la sociología pasaría a tener a su alcance cantidades de datos para describir y estudiar con detalle la evolución de las disciplinas científicas: pudiendo ayudar a

identificar tendencias y patrones en las publicaciones de revistas de impacto, lo que sería útil para los investigadores que buscan orientar sus propias investigaciones. (Martínez-Uribe, 2022).

### ***5.7 El derecho y el enfoque jurídico garantizando la legalidad de los procesos investigativos a través de big data***

El uso de macrodatos en derecho puede ser tan amplio como las peculiaridades humanas. Las diferentes materias y la propia diversidad de los sistemas jurídicos y legales hacen que la aplicabilidad de tecnologías big data sea bastante expresiva. Sin embargo, la mayoría de los artículos investigados se centran en la protección de los derechos fundamentales de las personas y agentes jurídicos. Un ejemplo es el texto de Garrido (2019), uno de los primeros que discute big data y derecho. A través de los ejemplos sobre el uso de GPS, el autor defiende la necesidad de mayor regulación y protección de los datos personales de las personas involucradas con datos de georeferenciamiento.

Un segundo texto demuestra la utilización de tecnología big data en investigación clínica y de salud pública como una interesante herramienta para la regulación y garantías jurídicas de las decisiones algorítmicas que están siendo implementadas en hospitales y unidades de asistencia a la salud (Hernández, 2020). La fiscalización de grandes conjuntos de datos financieros y transacciones relacionados a la salud, como compra de bonos, contratación de planes, indicación y compra de medicamentos, agendamientos de consultas y exámenes, son yuxtapuestas y analizadas con la finalidad de detectar patrones sospechosos o actividades ilegales como fraudes o compra de servicios médicos ilegales (Pérez, 2015). Los algoritmos y modelos de detección de anomalías pueden identificar, entonces, transacciones o comportamientos inusuales, lo que permite a las autoridades sanitarias intervenir y tomar medidas apropiadas, minimizando el impacto en fondos públicos.

En general, se destaca que el uso de macrodatos en el ámbito legal puede tener grandes ventajas, pero también puede amenazar los derechos de las personas, especialmente en lo que respecta a la privacidad (Téllez Carvajal, 2020). Por lo tanto, es necesario contar con las medidas adecuadas de control de la información, procedimientos transparentes y seguros que garanticen el máximo nivel de confidencialidad y el respeto a los derechos y libertades de las personas (Jara Fuentealba, 2021). Es importante destacar que el uso del big data en derecho plantea desafíos éticos y legales, con respecto a la privacidad, la transparencia, el manejo de datos y la protección de datos personales. La recopilación y el análisis de grandes cantidades

de datos requieren un manejo cuidadoso para garantizar el cumplimiento de las leyes y normativas aplicables a la realidad de cada país (Devins, et al., 2017).

### ***5.8 Ciencia política, opinión pública y participación ciudadana impulsadas por herramientas big data***

En el ámbito de las ciencias políticas, el uso de los datos masivos ha permitido a los investigadores y analistas políticos examinar y comprender a partir desde otras perspectivas, diversos aspectos de la política, como la opinión pública, la participación electoral, los procesos legislativos y la toma de decisiones gubernamentales.

Algunas de las implementaciones más comunes encontradas en mi primera búsqueda son de análisis de opinión pública (Arcila-Calderón et al., 2017) y evidencian que, como las redes sociales, los blogs y otros medios digitales, generan una gran cantidad de datos que pueden ser analizados para obtener información sobre las preferencias y actitudes de los ciudadanos. Estos datos pueden usarse para medir la opinión pública en tiempo real, identificar tendencias y predecir comportamientos electorales.

Esta utilización, cada vez más incorporada en los escenarios políticos y electorales, se vincula con una segunda implementación de big data en ciencias políticas: la predicción electoral. Big data puede utilizarse para predecir resultados electorales al analizar datos demográficos, datos de encuestas y otros indicadores (Shorey y Howard, 2016). Así, los científicos políticos pueden construir modelos predictivos que estimen los resultados de una elección, identificando patrones y factores clave que influyen en los resultados.

Por otra parte, ofrece la posibilidad de entender mejor la segmentación electoral, es decir, cómo funciona la fragmentación de los votantes en grupos más específicos (Nickerson y Rogers, 2014). Esto ayuda a comprender mejor las preferencias y necesidades de cada grupo de electores, permitiendo a los candidatos y partidos adaptar sus estrategias de campaña y mensajes para alcanzar a audiencias específicas.

Otra implementación relacionada a las ya presentadas, es la emergencia de estudios políticos basados en evidencias, permitiendo a los formuladores de políticas tomar decisiones más informadas a partir del análisis de grandes volúmenes de datos. Cuanto más complejos y más robustos los datos, mayores son las posibilidades de identificar nuevos patrones y tendencias

que ayuden a comprender los desafíos políticos contemporáneos, tanto global como localmente (Monroe, 2013). Por fin, el texto sostiene que la evaluación del impacto de las políticas implementadas y desarrolladas en base a datos masivos, digitales y georeferenciados, contribuyen para la implementación de políticas más efectivas y eficientes.

Por último está la vigilancia, detección de anomalías y la optimización de los procesos de responsabilidad política: big data también pueden usarse para monitorear las actividades políticas y detectar el desvío de funciones, recursos y propósitos políticos, como fraude electoral o corrupción (López Espinosa, 2022). Al analizar datos de múltiples fuentes, como registros electorales, datos financieros y medios de comunicación, acordes y textos políticos, se pueden identificar patrones sospechosos que requieran una investigación más profunda, que una vez averiguadas, posibilitan el entendimiento de los factores más relevantes del mal uso de los recursos públicos.

### ***5.9 Geografía social y económica: remodelando los espacios geográficos a través de big data***

En geografía, big data se usa principalmente para analizar grandes cantidades de datos geoespaciales. En general, los textos no dejan de remarcar la importancia de esta nueva posibilidad de extraer y utilizar una gran cantidad de datos georeferenciados, que antes no podían ser procesados conjuntamente.

La posibilidad de analizar una mayor diversidad de información geográfica y cruzarlas entre sí, es lo que permite a los investigadores identificar nuevos patrones y tendencias en la geografía (Liu et al., 2016). Junto a esto, está la ventaja de utilizar big data para la visualización de datos, espacios y territorios, cada vez más grandes y complejos.

Con respecto a la construcción del saber científico geográfico, se mencionan algunas áreas potenciales de investigación, como la integración de datos geoespaciales y no geoespaciales, la mejora de la calidad de los datos, el desarrollo de nuevas técnicas de visualización y análisis, y la exploración del potencial del aprendizaje automático y la inteligencia artificial, sin dejar de abordar los problemas éticos y legales relacionados al uso de estos datos (Kitchin, 2013).

Otra característica comúnmente mencionada en los textos de geografía, apunta a la capacidad de geolocalización. Se estima que el 80% de los datos masivos son espaciales, ya sea porque se dispone de las coordenadas o la dirección postal del lugar o porque el propio contenido de los datos hace referencia explícita a lugares concretos (Puebla, 2018). Por lo tanto, son datos cartografiables. Uno de los ejemplos del uso de tecnologías big data en su dimensión territorial son el análisis de datos procedentes de sensores para controlar la calidad del aire y el agua, la predicción de patrones meteorológicos y el análisis de grandes conjuntos de datos para mejorar la eficiencia y precisión de los sistemas de información geográfica (SIG) (Lansley, et al., 2019), muy útil para proveer información relevante sobre el clima, las condiciones de vida de la flora y fauna, la apropiación del territorio natural y el monitoreo ambiental de todo tipo.

Pero el trabajo del investigador de ciencias sociales con enfoque geolocalizado debe ir más allá del dato georeferenciado y de la generación de mapas (por ejemplo, mapas de *tweets*). Es necesario que los diversos profesionales en geografía estén atentos a la segunda fase del juego: pensar el “cómo” y analizar patrones y procesos espaciales mediante el empleo del análisis multiinformacional. Situar los datos en su contexto social, económico y político, como por ejemplo, los estudios de la geografía del consumo, son esenciales en el desarrollo de la geografía en tiempos de procesamiento computacional de información (Puebla, 2018).

### ***5.10 Periodismo y comunicación: las implicaciones de big data frente a la información digital***

La utilización de los datos masivos en periodismo y comunicación puede tener algunas ventajas y ser abordada desde diversas perspectivas. El periodismo de datos es un ejemplo de cómo la inmersión en una cultura de datos digitales puede ser aplicada a noticias y otros periódicos (Vivar, 2018). Al parecer, el big data permite a periodistas y comunicadores sociales recopilar grandes cantidades de información desde diversas fuentes, como redes sociales, registros gubernamentales y bases de datos públicos, que luego usarían para investigar y analizar tendencias, patrones, descubrir historias ocultas y generar contenido informativo para la población (Castillo-Esparcia et al., 2021).

Además, las capacidades de big data pueden impulsar la innovación empresarial y mejorar el rendimiento de las organizaciones por medio de la obtención de información detallada sobre el perfil de su audiencia. A partir de técnicas simples de extracción de datos en internet, una

empresa de comunicación es capaz de identificar preferencias, intereses, comportamientos en línea, patrones de consumo y medios de sus consumidores y usar esta información a su favor. Esto les ayuda a comprender mejor a su público objetivo y adaptar su contenido para maximizar su impacto, fortaleciendo así, su audiencia y sus intereses (Flores Vivar, 2018).

Sin embargo, la utilización de estos datos masivos también puede presentar limitaciones y riesgos, como la necesidad de preprocesamiento de datos y la gestión de información, autenticidad de los datos, preservación del registro, anonimización de colaboradores, en fin, temas vinculados a la logística de la utilización responsable, transparente y ética de los datos en y por la imprenta, está en juego en esta discusión, según nos alerta Wang, et al, 2016). En el ámbito del periodismo, la adaptación de los códigos de ética periodística a Internet y las TIC es un tema que cobra aún mayor importancia debido a los nuevos problemas éticos que han surgido en el ecosistema digital (Wiencierz y Röttger, 2017).

\* \* \*

A partir de la lectura inicial de los textos que componen las últimas sesiones, podemos ver que las implementaciones de big data a las ciencias sociales son innumerables. Prácticamente todos los textos evidencian un tipo distinto de uso de las herramientas de procesamiento de grandes datos y si bien cada artículo tiene su entrada y un análisis específico, pueden extraerse algunas conclusiones en común.

Los textos más antiguos encontrados sobre el tema datan de 2011 (Boyd y Crawford, 2011), y a partir del año 2015 esta muestra observa un aumento, lo que evidencia una cierta juventud o emergencia de los textos académicos que versan sobre big data en el marco de las ciencias sociales.

La mayoría de ellos están, en alguna medida, involucrados con la discusión metodológica y aspectos disciplinarios por la cual la introducción de la manera big data de trabajar pueden traer al interior de sus campos de estudio. Los retos metodológicos resaltados por los textos a partir del uso de big data en las distintas disciplinas en ciencias sociales son varios.

Destaco la preocupación con el proceso de tratamiento de datos previo a su uso, como una advertencia general presente en los textos. La necesidad de reducir la dimensionalidad de los datos para poder analizarlos de manera efectiva, y la selección de las variables que más

aportan en términos de información útil, son ejemplo de discusión que involucra aspectos de capital humano cualificado para tal labor, y representa una dimensión faltante a las investigaciones de esta naturaleza, un consenso general y presente en los textos estudiados.

Otro reto importante, evidenciado en la mayoría de los artículos, es garantizar la calidad y fiabilidad de los datos, ya que pueden contener errores o sesgos que afecten los resultados del análisis. Además, es necesario tener en cuenta las limitaciones éticas y legales en el uso de datos personales y datos sensibles. La sociología, el derecho, la historia y la geografía social, son las disciplinas que más textos aparecen, en mi búsqueda, problematizando esta cuestión.

Por último, hay en los textos una clara mención al papel preponderante del desarrollo de nuevas herramientas y técnicas para analizar grandes cantidades de datos, lo que requiere una combinación de habilidades técnicas y conocimientos disciplinarios, haciendo con que los textos poseen una ligera tendencia a problematizaciones de cuño metodológico. Sin embargo, la gran mayoría de estos textos no provienen de resultados y hallazgos de investigaciones aplicadas en/con big data tienden a ser más teóricos, problematizando a nivel epistemológico las metodologías posibles de investigación a partir de técnicas big data.

En términos de contenidos de fondo, es bastante evidente que los textos analizados en este primer acercamiento poseen un tono de problemas futuros. Más que demostraciones o evidencias de las investigaciones aplicadas en su contexto de investigación o trabajo académico, los artículos buscan abordar teóricamente las implementaciones o posibles efectos que el uso de los macrodatos puede tener en sus respectivas disciplinas. En este momento, al parecer, los científicos están abriendo / liderando un terreno más conceptual y descriptivo de lo que se podría hacer a partir de la tecnología big data. Cuando mucho, están ejemplificando investigaciones aplicadas de tercero o citando experiencias y textos de otros investigadores.

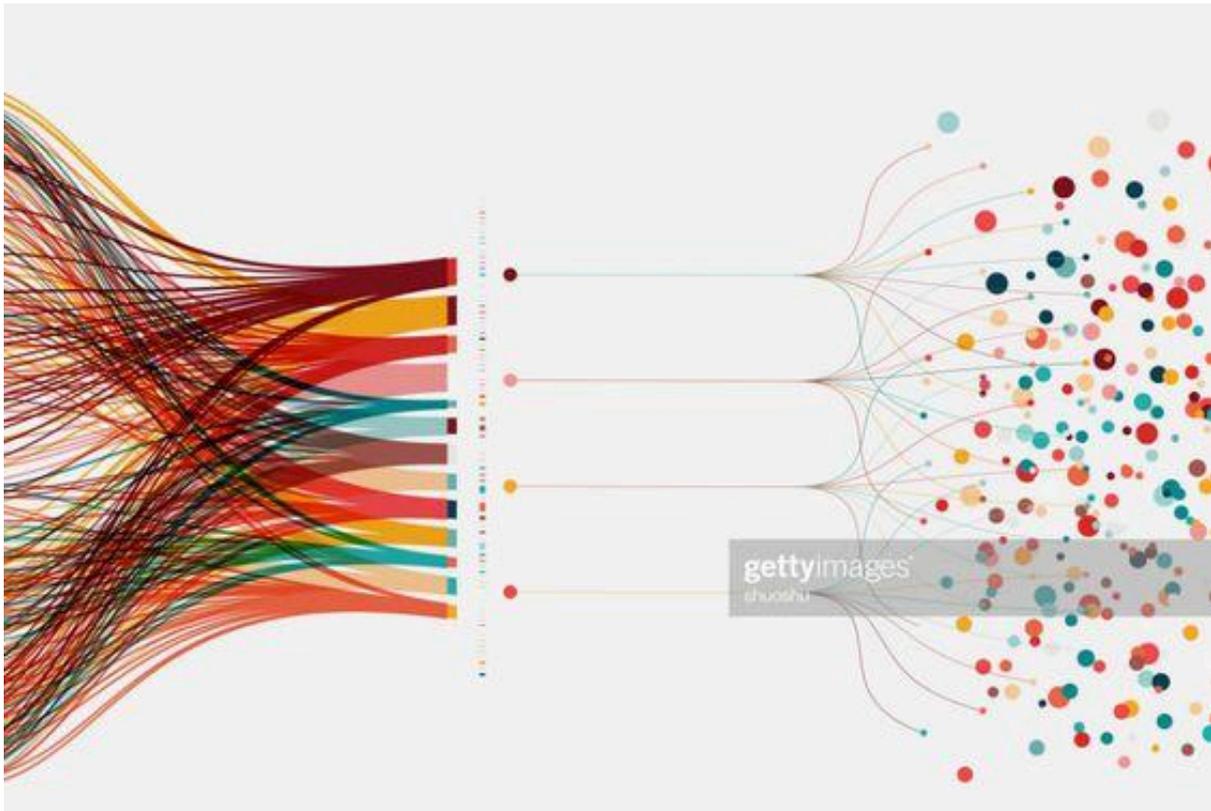
En lo que se refiere a la lectura de dichos textos y mi trayectoria personal de investigación, no es menor mencionar que desde un principio supe que la conducción de una etapa bibliográfica de investigación no sería suficiente para entender los matices de trabajar computacionalmente los datos digitales sobre la sociedad. La inquietud por comprender lo que estaba siendo hecho y aplicado en las investigaciones en torno a big data, me abrió la oportunidad de trabajar ejerciendo la función que estudiaba, lo que me permitió aún mayor inserción en el campo de estudio. Así, mientras exploraba los documentos científicos,

emergía la segunda etapa de mi investigación: la etnografía de laboratorio fluyó sin grandes esfuerzos, y los entrevistados fueron llegando también gracias a mi contacto con el campo.

Las lecturas también me permitieron comprender el grado de interrelación que estas disciplinas tienen entre sí, además del cruce con disciplinas de ciencias exactas e ingenierías. Aunque abordan distintos problemas, por lo general, la manera en que utilizan las técnicas de procesamiento de información son muy similares entre sí, sobre todo tratándose de procesamientos de lenguaje natural y texto. Esto ayuda a entender el por qué se requiere el constante tránsito entre disciplinas, vocabularios y habilidades de traducir e interpretar informaciones de diversas naturalezas.

Finalmente, mencionar que en los anexos se puede encontrar una tabla con las referencias bibliográficas de los 5 textos principales mencionados hasta acá, distribuidos según su disciplina, ordenados por título y palabras claves, en esta fase inicial de diseño de investigación, para que el lector pueda consultar bibliografías y enfoques específicos en caso de interesarse, o para quien busque profundizar en sus estudios de ciencias sociales y big data.

## PARTE 3: CAMPO



“Igual que el etnógrafo en una cultura extranjera, el científico en el laboratorio se enfrenta con ruidos y con ilimitadas incertidumbres de las cuales extrae sentido, valiéndose de conceptos y de procedimientos que por el momento son aceptados. Como en la etnografía, las incertidumbres relevantes aparecen aquí en el nivel de reconocimiento, de la identificación y de la atribución de un sentido a los datos y las observaciones”. (Knorr-Cetina, 2005, p.313)

## Cap 6: Ruta etnográfica: contexto de investigación

Bienvenidos a la parte que más me gusta de todo el proceso de investigación. Siento mucha pasión por la experiencia en terreno. Es una especie de viaje para mi: lo que más amo hacer en la vida. Me encuentro otra vez, escribiendo informes de campo, un relato hecho tesis de mi más reciente vivencia etnográfica. Aunque en mi carrera como profesional de científica social he hecho innumerables trabajos etnográficos en diversas ocasiones, para diferentes investigaciones, este resultó ser bastante diferente. Es la primera vez que he profundizado en los estudios de la ciencia, de la tecnología y su vínculo con la sociedad y también es la primera vez que lo hago en un ambiente totalmente digital y online.

La pandemia mundial COVID-19 impactó el mundo, y entre otras muchas consecuencias de este periodo, nos vimos sin más alternativa que conducir gran parte de los procesos etnográficos e investigaciones que involucran algún tipo de terreno, acontecer de manera virtual. Los científicos sociales la llamamos etnografía digital<sup>56</sup> (Hine, 2015; Bárcenas y Preza, 2019). Tal situación, para muchos de nosotros, configuró ser una experiencia muy desalentadora: Conducir una etnografía digital tiene muchos desafíos y hay que mencionar que la capacidad de percepción que sólo un cuerpo atento puede realizar, se ve disminuida, por ejemplo. Más que esto, las sutilezas de los no dichos, las miradas y señales entre las personas, suelen ser aún más evidenciadas en la presencialidad. No es fácil captar los sentidos más allá del habla, en un espacio de reunión donde la mayoría de las personas están con las cámaras apagadas, lo cual le cuesta carísimo al proceso etnográfico.

Por otro lado, sabemos que, con relación a varios otros contextos, la digitalidad mejoró la experiencia del *que hacer* etnográfico (Salmons, 2017) posibilitando nuevas herramientas para captar y entender las nuevas prácticas sociales contemporáneas, y en contexto digital, entendiéndolas como reales en su experiencia y en sus consecuencias (Hine, 2015).

Afortunadamente, a lo largo del proceso etnográfico, mi inserción en el terreno me mostró que aún en contexto de digitalidad era posible entender algunas dinámicas, así como la estructura por detrás de las investigaciones aplicadas en big data. Por lo menos en el

---

<sup>56</sup>Entendiendo etnografía digital como una adaptación de los métodos tradicionales de investigación antropológica a los nuevos medios. Actualmente, se considera la metodología apropiada para proporcionar una comprensión sistemática de diversos fenómenos en Internet, a vida digital y las prácticas cotidianas desplegadas por los sujetos –en mediación con la tecnología– en la esfera en línea (Hine, 2015; Bárcenas y Preza, 2019). Posicionarse desde esta opción metodológica implica asumir, desde lo epistemológico, que Internet y lo digital son experiencias cotidianas, integradas y encarnadas (Hine, 2015) y que, por tanto, la dimensión en línea (*online*) como la fuera de línea (*offline*) constituyen parte integral del entramado las nuevas prácticas sociales propiciadas y desplegadas en Internet.

laboratorio en el que yo me propuse estudiar. También se configuraba como la única manera de realizar el estudio, ya que todo el laboratorio se reunía virtualmente a través de la plataforma de videoconferencia para realización de reuniones, así como, individualmente en sus casas, tras el computador, trabajaban en sus funciones de investigación y para los proyectos en que participaban.

Conducir una etnografía digital también significó más flexibilidad con respecto a la recolección de información y documentación del campo. Con la masiva emergencia de las videoconferencias, todas las reuniones, congresos, ponencias, comenzaron a quedar grabadas en sus canales de Youtube o página web de los eventos. Esta realidad me permitió, por ejemplo, participar en diferentes momentos de congresos que acontecían simultáneamente, o, por el contrario, no participar *en vivo* de un seminario sobre el tema, pero tener la chance de verlo por grabación, un día después, con subtítulos en español, y mejorando muchísimo mi comprensión en una ponencia hecha en francés o inglés.

La pandemia global que asoló el planeta a partir de los años 2019, adelantó la digitalidad en diversos frentes, servicios y costumbres sociales. Vimos nacer aplicaciones, sistemas y recursos que hasta antes de este evento, no habíamos experimentado. Durante por lo menos 12 meses, el teletrabajo fue la opción de la mayoría de los profesionales y todos nos vimos interpelados a adaptarnos. La belleza de las crisis, es que si miradas con atención, nos pueden proporcionar tremendas oportunidades para inventar o incorporar nuevas maneras de hacer lo mismo: recolectar información de calidad. Nuevas maneras de evidenciar lo que vas encontrando en las hablas y discursos de los sujetos investigados. Y eso fue lo que pasó conmigo al final.

Escuchábamos de colegas químicos, psicólogos y biólogos que habían perdido años de investigación por no poder ir al laboratorio a alimentar sus bacterias, proteínas y ratones, por las restricciones de movilidad. Mi problema más alarmante era tener que adaptar entrevistas y observaciones a la realidad virtual, y hacer todo a través de la pantalla de un computador. Dificultad, que al lado de estos colegas, no representaba un gran desafío.

Considero que estudiar temas de tecnología, digitalidad y datos masivos, en un momento donde todo el mundo vivía su vida a través de estos soportes, se hizo especialmente interesante. Muchas oportunidades de estudio de las temáticas digitales se volvieron aún más relevantes. Los financiamientos en 2020/2021 auspiciaron algunas importantes investigaciones e iniciativas con respecto a políticas de datos, procesamiento computacional

de información estratégica y seguridad, transparencia y buen uso de datos públicos en América latina. Hemos avanzado, aunque a pasos lentos, con respecto a las cuestiones legales y reglamento de la políticas de datos en muchos países de latinoamérica<sup>57</sup>.

Estudiando temas afines a la digitalidad y tecnología, y por estar insertada en este medio, quedé estratégicamente bien posicionada para ocupar unos de los puestos de investigación que se atribuían a estos proyectos. El desarrollo de esta etapa de la investigación, la fase etnográfica, sólo fue posible gracias a una compañera de mi generación en el doctorado, que me envió un correo que le había llegado con una oferta de contratación que “era hecha para mi” (en sus palabras), y de verdad lo fue. No podía creer cuando leí la descripción del puesto de trabajo. La cereza del pastel era el nombre del cargo *etnógrafo en ciencias de datos*.

Después de pasar por un intenso proceso selectivo quedé contratada y fue así que las puertas del campo de investigación se abrieron con un pase de magia. O como diría Knorr-Cetina, gracias a las *conexiones transcientíficas* de la investigación.<sup>58</sup> Así que los resultados, las discusiones y el compartir de la construcción del conocimiento científico sobre las ciencias sociales contemporáneas, digitadas por estas manos, es fruto de una vivencia como científica social, que trabaja e investiga la temática big data en el ejercicio de sus actividades laborales y académicas.

Como he mencionado en el transcurso de este estudio, la fabricación de los saberes acá presentados es parte intrínseca de una búsqueda por encontrar respuestas al desarrollo de mi propia profesión, mientras me educó para ello. Es por eso que acredito y defiendo con seguridad, la fortaleza de una investigación implicada. No es sin motivo que tengo a mi lado la conversación con Knorr-Cetina y los constructivistas de los estudios de la ciencia, tecnología y sociedad (Mulkay, 1979; Callon, 1982, 1986); Lynch, 1983); Latour, 1995, 2005, 2011).

---

<sup>57</sup>Durante una reunión de balance final del año, la dirigente de la reunión mostró el número de iniciativas, proyectos y premios que el laboratorio había adquirido. Según ella, 2021 había sido particularmente especial para el laboratorio debido a la cantidad de proyectos y esfuerzos que se pudieron asociar. “La gente está interesada en la ciencias de datos y ahora está comenzando a entender su valor, importancia y cómo funciona esta cuestión. Así que estamos medio que surféando una ola muy buena y estoy orgullosa porque felizmente, gracias al equipo y la buena voluntad de todos ustedes, hemos conseguido surfear” (Reunión cierre del año, Diciembre, de 2021)

<sup>58</sup> Knorr-Cetina, en la parte 9 de su libro, argumenta que estas condiciones son determinantes para la innovación y desarrollo de la producción del conocimiento científico. La transciencia es un concepto que admite al interior de la fabricación de los saberes un movimiento dinámico y que combina entre sí los diferentes e inesperados (o sorprendentes) tipos de recursos, preguntas, contextos y esfuerzos deliberados por parte de los científicos y de aquellos que están involucrados financieramente (los auspiciadores).

Mi cercanía e involucramiento en el campo y estar al frente de la investigación de la misma naturaleza de mis objetivos de estudio doctoral, me permitió acceder a reuniones, entrevistas, personas y espacios, que de otra manera no hubiera sido posible. Además, abrió puertas para consolidar una red de investigadores y trabajadores sociales latinoamericanos en análisis de datos digitales y masivos, fortaleciendo así el vínculo entre los interesados en la discusión del tema en la región. Conformar diversos paneles y mesas de trabajos en congresos de CTS internacionales, trabajando en pro de la difusión del conocimiento big data en las ciencias sociales. Actualmente, espero estar abriendo caminos conceptuales para los colegas que se vean interesados en conocer más de cerca la consolidación de un campo disciplinar social en emergencia y en pleno proceso de expansión.

La fortaleza de este estudio es justamente el hecho de que las *conexiones transcientíficas* de mi actividad académica dieron respaldo y soporte para ejercer mi trabajo de investigadora social en *big data/ciencias de datos* de forma más bien instruida y preparada. Tanto desde el punto de vista teórico, donde construía analíticamente las bases conceptuales para mi estudio, como también desde el punto de vista práctico, por estar insertada en el mismo campo profesional de estudio. “Estar metida en el contexto”, parafraseando a mi profesora del taller de tesis, y, así, poder vivenciar los desafíos y posibilidades de trabajar en un contexto de investigación en ciencia de datos, me proporcionó tener contacto con la realidad, de la vivencia interna con investigadores asociados, que ninguna lectura, por más esclarecedora que fuera, me podría haber proporcionado.

Lo que le presentaré en la secuencia, es por lo tanto, una colección de hallazgos proveniente de 1 año y 4 meses de estudios en terreno, como investigadora y trabajadora en ciencias sociales interesada en la temática big data, entre agosto de 2021 a diciembre de 2022 en el laboratorio GobLab UAI: Laboratorio de innovación pública de la Escuela de Gobierno de la Universidad Adolfo Ibáñez, en Santiago, Chile.

En este capítulo les presentaré las estrategias de campo utilizadas, bien como les contaré un poco sobre mi terreno. En los que siguen (7, 8, 9 y 10) les presentaré los principales hallazgos provenientes de la etnografía emprendida en el laboratorio que trabajé e investigué. Son agrupamiento de conversaciones durante la realización de proyectos, por entrevistas semi-estructuradas, conversa entre colegas en las reuniones de trabajo, o participaciones en alguna actividad relacionada (como congresos, presentaciones del laboratorio, talleres,

seminarios y ferias organizados por diversas entidades y organizaciones alrededor de Chile y Brasil, principalmente.

### **6.1 La etnografía sensitiva y digital como enfoque etnográfico para los estudios de laboratorio involucrados en la temática big data**

La idea de emprender una etnografía siempre estuvo presente en mi interés de investigación. Primero, por haber demasiasdas buenos estudios de laboratorio al interior de las CTS mundiales: Latour y Woolgar (1979), Callon (1986), Knorr-Cetina (1981), Lynch (1982), Bourdieu (1993), habían hechos investigaciones excepcionales habiendo entonces, influenciado y convencido a toda comunidad académica, sobretodo las de ciencias sociales, de la importancia de estudiar el laboratorio como un espacio de construcción y fabricación de conocimiento por los científicos.

Pensar en conducir un estudio etnográfico en un laboratorio con estas características, surge casi naturalmente, aún más dada mi inserción laboral durante este proceso de investigación. Las dimensiones que serán presentados de ahora en adelante son resultado de un proceso de investigación *etnográfica digital* y la utilizacion de la metodología *sensitiva*.

Utilizo la metodología sensitiva (Knorr-Cetina, 2005) como soporte de mi investigación porque creo en la práctica que se desarrolla por detrás de su raciocinio conceptual defendido en esta manera de recolectar informaciones en campo. Hay por lo menos 3 características fundamentales en su manera de producir etnografía e información relevante para los estudios de laboratorio que se adecuan satisfactoriamente a los objetivos de esta investigación.

Antes de apuntarlas, se hace necesario recordar que, como apunta Knorr-Cetina (2005), que el trabajo científico por detrás de cada publicación (texto o documento científico) se encuentra y se manifiesta en la vida cotidiana de un laboratorio. Percibir las representaciones sobre lo social imbricada en la escritura de los artículos académicos analizados, significa también conocer las prácticas científicas por detrás de cada texto, cada escrito, cada documento. Conducir un estudio etnográfico en un laboratorio que trabaja, produce y escribe sobre big data y ciencias de datos, fue una decisión que tomé al emprender este proyecto, objetivando comprender las dinámicas contextuales presentes más allá del texto.

Otro por menor antes de adentrar a los porqués de elegir la metodología sensitiva como marco conceptual que guía y da forma a mi entrada a terreno en este trabajo: es necesario explicar que aunque las condiciones de estudio de la autora, bien como el contexto investigativo tanto no que dice respecto a la época en que conduzco su propio estudio (finales de los años 70), y el tipo de laboratorio que estudiaba (laboratório de biologia molecular) són bastantes diferente a los míos, defendo que tales diferencias no resultan en una menor o errónea abordagem conceptual para enmarcar mi trabajo de investigación. Primero por que esta investigación, tal cual la autora, me propongo a observar las prácticas diarias, las interacciones sociales en los métodos de trabajo de dos científicos para, así entender cómo sus conocimientos son producidos y legitimados al interior de un laboratorio, y por consecuencia es traspasado a determinados aspectos de la sociedad. Este trabajo minucioso de comprensión de las dinámicas laborales y su consecuencia con relación a difusión del conocimiento científico en un campo, puede ser apreciado, **en este trabajo a partir del capítulo 8 (Investigaciones en un laboratorio en ciencias de datos).**

Tal como Knorr Cetina (2005) los análisis hechos focalizaron en las dinámicas internas del laboratorio visando explorar cómo el conocimiento científico no es simplemente un descubrimiento de hechos naturales, sino algo que se “fabrica” a través de procesos sociales, materiales y técnicos presentes en los laboratorios, que a su vez, utilizan recursos y soportes adecuados a un momento y condiciones sociales dispuestas localmente. Así como la autora, analizó el papel de los instrumentos y de los procedimientos experimentales, demostrando al fin, que el conocimiento científico es construido activamente por la conjunción de científicos, técnicas, recurso y contexto social absorbidos, siendo entonces, los “nuevos” conocimientos producidos no meramente “descubiertos”.

En la ausencia de un estudio profundo y consolidado similar a temática trabajada en esta investigación de los demás autores consagrados de la CTS, lo que se justifica por la novedad y emergencia de mi objetos y sujetos de estudio, defiende la utilización del modelos de investigación de análisis de la autora no solamente apropiado, sino que suficiente para dar conta de los aspectos y matices que he estudiado en dicha investigación.

Paréntesis hecho, volvamos a la primera de las características que me impulsaron a elegir esta corriente etnográfica en especial: su vehemente reiteración de la necesidad de guiar una cercanía intersubjetiva con el campo de trabajo.

Un enfoque sensitivo de estudio de la ciencia, como señalé antes, nos obliga a desechar los intermediarios metodológicos generalmente usados para la recolección de datos. Debemos renunciar a los servicios de entrevistadores, cuestionarios y oficinas de estadísticas, y exponernos nosotros mismos, mediante la observación directa y la participación, al significado salvaje de la acción de esos científicos de laboratorio. [...] Si estamos interesados en la producción y en la reproducción de los hechos científicos (que, como Whitley ha señalado tan adecuadamente, es todavía una caja negra para los estudios sociales de la ciencia) podemos estar bien avisados si buscamos la intersubjetividad dejando fuera el embrague metodológico, o sea, mediante el uso de la observación cercana del lugar de producción (Knorr-Cetina, 2025, p.104)

Mi explícito involucramiento con el campo, imprime novedad e información privilegiada en cuanto al proceso de construcción de los saberes sociales emprendidos alrededor de los macrodatos y el procesamiento computacional de estos en las ciencias sociales.

El segundo principio está íntimamente conectado al primero. Si es necesario producir una cercanía con los sujetos y espacios observados, también es importante dejar que el campo hable. Más que esto, la metodología sensitiva, por medio de su principio número dos, tiene como objetivo la atención a los diseños metodológicos y a técnicas de producción de información que puedan captar la voz de aquello de lo que habla. Lo que nos falta en la antropología no es “la falta de comprensión, sino que controlar la constitución conceptual dada a esa comprensión en la representación o en la transmisión de los conocimientos” (Knorr-Cetina, 2005, p. 91).

Confirmando, entonces, la importancia de haberme hecho parte como trabajadora e investigadora en los proyectos pertenecientes al laboratorio que me propuse estudiar. Acceder tan libremente a una red de instituciones, organizaciones, investigadores y académicos interesados en investigaciones con enfoque público y social en un laboratorio de Latinoamérica, facilitó mis capacidades de poder escuchar no sólo a los científicos autores de los artículos estudiados, sino también a decenas de compañeros, trabajadores de big data en diversas instancias académicas y profesionales. Esto me permitió estar atenta y aprender más

fluidamente el “idioma computacional”, y a su vez, escuchar y aprender las sutilezas presentes en la construcción de saberes institucionalizados frente a situaciones como negociaciones de proyectos, promoción y venta de diplomados, construcción de seminarios, reuniones de contrataciones de asistentes de investigaciones y académicos de pasantía; formación de colaboración entre los más variados científicos sociales o no, interesados en el análisis de datos. Grandes datos.

El tercer principio del enfoque sensitivo es muy clave: la atención a las prácticas. La mirada atenta a los que se hace, a los verbos de las prácticas producidas al interior de los laboratorios y sus proyectos. “El tercer rasgo distintivo a especificar aquí es el *interaccionismo metodológico*, que garantiza que esta etnografía se interese en la práctica, más que en la cognición, de sus sujetos”. (Knoor Cetina, 2005, p.92).

Conforme profundizaba en los estudios de la producción escrita y académica de los científicos sociales involucrados en temáticas de datos masivos en función de conocer lo social, este principio me ayudó a esclarecer la necesidad de conversar con los científicos y autores pertenecientes a distintos laboratorios alrededor de Latino América.

La conducción de las entrevistas a los autores de decenas de publicaciones fueron una manera de dar voz a los sujetos involucrados en este estudio, como también tener la oportunidad de a partir de estos diálogos y momentos de interacción, percibir —a partir de sus hablas y mis propias percepción, considerando la lectura de sus textos— las prácticas, artesanías, mecanismos y negociaciones por detrás del conocimiento que buscaban evidenciar o difundir por medio de sus artículos (Lo que será mejor discutido en el capítulo 9 de este trabajo).

La digitalidad fue fundamental para la amplitud de este estudio. La convencionalidad y la masividad de poder hacer una reunión *online* y de cualidad en cualquier parte del mundo, facilitó las entrevistas e interacciones en congresos, seminarios y pequeñas instancias de formación, en diferentes países del mundo, que seguramente en modalidad presencial no sería posible sin un volumen alto de financiamiento por detrás mio.

La digitalidad del mundo del trabajo y de muchas investigaciones, me posibilitó con mayor facilidad obtener información por las cuales salieron la representación de las controversias: los disensos, los discursos más dominantes, las estrategias más utilizadas, las áreas más desarrolladas en los estudios utilizando big data, fue posibilitada por la capacidad de acceder

a varios espacios virtuales dedicados a mis propios objetivos, o adyacentes a la problemática de grandes datos, sin mayores costos, inclusive y sobretodo financieros.

Es justamente bajo este enfoque que se cruzan, entonces, la mirada de la antropología sensitiva con la etnografía digital —defendida por Christine Hine en su libro *Ethnography for the Internet: Embedded, Embodied and Everyday*, en 2015.

La *etnografía digital* implica la inmersión del investigador en comunidades en línea, redes sociales y otros espacios digitales (Hine, 2015). A través de la participación activa y la observación, el investigador busca comprender cómo las personas se comportan, interactúan y crean significado en estos entornos.

Yo no estaba interesada específicamente en el relato de la vida en línea y todo el proceso por lo cual vivimos la digitalización de la información y producción masiva de datos, pero acabé siendo interpelada a incluir esta dimensión en mi observación y métodos de obtención de información. No es posible hablar de big data en las investigaciones sociales, sin considerar el contexto de digitalidad, ni tampoco el entorno global que vivimos. Acompañar de cerca mi comunidad de estudio significó imergir en una vida completamente asociada a las pantallas y ambientes fuera de la presencialidad.

La etnografía digital no es simplemente una extensión de la etnografía tradicional, sino que representa un enfoque específico para investigar la vida en línea y cómo las personas interactúan en el entorno digital. Introducir estos principios en el estudio que se viene presentando, significó también un bocado de adaptación y flexibilidad a los soportes y artefactos virtuales y plataformas nativas, necesario para se hacer presente y participativamente observante (Platt, 1983) en las instancias de convivencia y trabajo colectivo.

En el caso de mi experiencia con esta investigación las reuniones de equipo de trabajo de los diferentes proyectos en curso acontecían, en la mayoría de las veces, en aplicaciones de videoconferencias. Las demandas, las tareas individuales o grupales eran agendadas y puesta en seguimiento en plataformas a depender de la institución colaboradora o adyacente al proyecto.

El acceso a los materiales presentados a posteriori a las instancias virtuales, permitió a mi también recolectar fácilmente una cantidad significativa de materiales y documentos, que

también son usados como evidencias y herramientas de elucidación para el montaje de este gran rompecabezas final.

Los análisis de estos materiales fueron construidos en base a un proceso de *Revisión Documental* (Sampieri et al., 2003): una técnica de observación complementaria que permite construir ideas del desarrollo y de las características de procesos, informaciones y contenidos presentes en un corpus textual. Puede ser realizada por medio de cartas, actas, planillas, informes, libros, imágenes, folletos, manuscritos y materiales audiovisuales.

Los documentos bajo esta mirada, representan el registro de la historia escrita o grabada y pueden estar relacionadas a acciones, experiencias y maneras de concebir ciertos fenómenos, situaciones y temas (Sampieri et al., 2003). Se cree que, en paralelo al proceso de observación objetiva de las publicaciones académicas que tratan el big data en ciencias sociales, la revisión de los documentos se constituye como recurso material de evidencia respecto de las decisiones, acuerdos y consensos al interior de cada temática trabajada en las instancias de observación.

Al tener la opción de recolectar a través de pantallazos, la conservación de enlaces y archivos de presentaciones en congresos y seminarios, se tornó posible recolectar en primera fuente una diversidad de conceptos, contactos y contenido. Así también fue posible conseguir muchos emails de entrevistados, permisos de participación como observadora de instancia que serían cerradas al público. Todo este material documental fue siendo utilizado para profundizar la comprensión y posterior creación de los hallazgos que se presentaron a lo largo de la parte 3 del texto.

Como Hine (2005) enfatiza, la vida en línea no se limita a un sólo sitio web o plataforma. La etnografía digital se extiende a través de múltiples sitios y espacios digitales, ya que las personas pueden participar en diversas comunidades en línea y plataformas de medios sociales. Esta afirmativa también es válida para el contexto de producción de información y levantamiento de datos en investigaciones realizadas en contexto digital, requiriendo alto grado de adaptación y sentido de urgencia de los investigadores en/para/con temas asociados a la digitalidad.

Tal realidad puede ser encarada como una ventaja, pues la capacidad de participar y recoger información relevante para los estudios emprendidos crece exponencialmente; pero también

puede ser un atropello, si el científico no maneja bien la organización y categorización de su información, es decir la calidad de su base de datos.

Temas como la seguridad, privacidad y el buen uso de las informaciones levantadas, es otra dimensión esencial a ser considerada en trabajos e investigaciones de esta naturaleza. Lo que está en la red se detecta y asocia fácilmente a un persona física, lo que incide negativamente en los derechos en cuanto a la protección de los datos personales y sensibles de los sujetos involucrados (Lessig y Lawrenc, 2002).

Consciente de este desafío, y de mi responsabilidad como una científica insertada en este contexto, es que declaro haber tomado precauciones sobre la seguridad y confiabilidad de los datos y el nombre de los científicos entrevistados y partícipes del proceso etnográfico de este proyecto investigativo. El uso de los nombres completos, fue previamente autorizado por la persona, con la intención de prestar un homenaje a los involucrados que tanto aportaron para el proceso de consolidación de esta investigación.

Los demás sujetos, sobre todo los autores de los artículos de los estudios bibliográficos, fueron entrevistados y grabados con consentimiento previo, participando de manera voluntaria, según la disponibilidad de cada uno<sup>59</sup>. Las grabaciones fueron transcritas, algunas a través de la contratación de un servicio de transcripción profesional y debidamente certificado, otras fueron hechas a partir de softwares diseñadas para tal.

Una vez transcritas, las entrevistas se guardaron en un *drive* en la nube, y una copia de seguridad en mi disco duro externo personal, ambos protegidos a través de tecnicas de seguridad. Los nombres de los entrevistados fueron anonimizados.

Más que dar voz a los sujetos investigados a través de las entrevistas y de la etnografía de laboratorio, digital y sensitiva, esta etapa de investigación tiene como objetivo comprender los pormenores de las investigaciones llevadas a cabo bajo la temática planteada en el proyecto. Es la fase del estudio para dar a conocer los matices de la práctica investigativa al

---

<sup>59</sup> Uno de los autores entrevistados, estaba muy ocupado, por lo que luego de conversaciones por email, decidimos en conjunto que la mejor manera de que realizar la entrevista era vía Whatsapp. Conseguimos negociar un número de 15 preguntas, que fueron previamente enviadas al celular del científico, quien me respondió en audios de entre 2 a 4 minutos. Resultó ser un formato interesante y eficiente de entrevista. La transcripción es más fácil de ser realizada y las preguntas ganan respuestas más objetivas y mejor elaboradas. A parte que el entrevistado controla su tiempo de respuesta permitiendo una mayor reflexión sobre el asunto. Decidí incorporar esta “técnica” en dos otras dos entrevistas más, pues estábamos con dificultades para conciliar agendas. Notas del Diario de campo, noviembre de 2021.

interior de este laboratorio y las consecuencias relacionadas a su forma de investigar y el tipo de producción que se genera a partir de ellos.

Defiendo que la mejor forma de comprender la relación entre la producción de conocimiento social instaurada al interior de los grupos de investigación y la forma de investigar “bigdatamente”, está en el análisis de no solamente la producción académica de estos profesionales/instituciones, como ya lo hice por medio de los estudios bibliograficos presentados, pero tambien en la observación etnografica del *que hacer* laboral de estos mismo perfil de investigadores y/o organizaciones.

El objetivo central de este segundo momento de estudio es profundizar en los contenidos que se producen al interior de la construcción del pensamiento social latinoamericano, y los impactos o transformaciones que la práctica de investigar los fenómenos sociales por medio de big data y el paradigma computacional, generan al interior de la disciplina.

Analizar las representaciones (interpretaciones) de lo social contenidas en las publicaciones científica en Latinoamérica y que discuten la problemática big data en sus investigaciones, significa no sólo estar presente compartiendo experiencias con este grupo de interés, más también reconocer los laboratorios científicos como una dimension más de analisis de las practicas y dinámicas sociales presentes en la contemporanead.

Es por esto que la lectura de los artículos, vinculada a la experiencia de observar el trabajo de este perfil de científicos, posibilita especular y conjeturar sobre las prácticas sociales de la actualidad. Sociedad y científicos hacen parte de un contexto social que refleja interés de investigación, facilita obtención de recursos, lo que a la vez, valida los intereses, tanto de la sociedad, como del científico, en cuanto miembro de una comunidad, o fruto de determinaciones sociales.

Las secciones que siguen, intentan detallar la descripción del laboratorio de estudios, con el fin de demostrar cómo los oportunos del laboratorio, los esfuerzos en condensar razonamiento lógico a las prácticas y actividades ejercidas y la propia estructura de interés de parte de los científicos, se muestran de manera socialmente organizada. Es decir, “cómo la continuidad contextual de las selecciones de laboratorio se presenta al mismo tiempo como una contingencia social” (Knorr-Cetina, 2005, p.176).

Los capítulos que se siguen, apoyado en el esfuerzo etnográfico de investigación, objetiva hacer el puente entre la vida de un laboratorio de innovación pública y en ciencias de datos y interesado en la temática big data, entre las representaciones y demandas sociales expresadas a través de los estudios y la producción de conocimiento de estos mismos científicos. La idea es armar un correlato de la interdependencia y de la conexión social que hallamos en las relaciones de recursos que atraviesan los campos transcientíficos (Knorr-Cetina, 2005).

## 6.2 El laboratorio como herramienta de análisis social

¿Qué es, después de todo, un laboratorio? Una acumulación local de instrumentos y aparatos, dentro de un espacio de trabajo conformado por mesas y sillas. Cajones llenos de utensilios menores, repisas cargadas de productos químicos y recipientes de vidrio. Heladeras y congeladores llenos de muestras cuidadosamente etiquetadas y de materiales-fuente: soluciones pulidoras y hojas de alfalfa finamente picadas, proteínas de una sola célula, maestras de sangre de ratas de ensayo y lisozimas. Todos estos materiales-fuente han sido especialmente cultivados y selectivamente alimentados. La mayoría de las sustancias y de los productos químicos son purificados y han sido obtenidos de industrias que proveen a la actividad científica o de otros laboratorios. Pero hayan sido compradas o preparadas por los propios científicos, esas sustancias no son menos producto del esfuerzo humano que los aparatos de medición o los trabajos escritos que están sobre los escritorios. Parecería, entonces, que a la naturaleza no se la va a hallar en el laboratorio, a menos que se la defina desde un principio como producto de un trabajo científico. (Knorr-Cetina, 2005, p. 58)

La importancia de los estudios de laboratorio en el contexto de las Ciencias, Tecnologías y Sociedad (CTS) y la documentación de los procesos que intervienen en la producción del conocimiento científico ha sido objeto de un extenso debate entre los científicos sociales desde la década de 1970. Durante este período, destacados investigadores como Mulkay (1979), Callon (1982, 1986), Knorr-Cetina (1981, 1983), Lynch (1983) y Latour (1995, 2005, 2011) se propusieron estudiar el desarrollo de la ciencia, su interacción con la sociedad y la tecnología a través de etnografías de laboratorio. Estos estudios han transformado nuestra percepción y comprensión de la ciencia y la construcción del pensamiento científico, así como su impacto en el desarrollo de la sociedad.

La comprensión de la interacción entre la ciencia, la tecnología y la sociedad se ha convertido, entonces, en tema central de muchos estudios con este enfoque, como también es el caso de la investigación que acá se presenta. Es en este contexto que los estudios de laboratorio desempeñan un papel fundamental al proporcionar a los estudios de la ciencia, un escenario/territorio bastante específico y interesando en asuntos, experiencias y investigaciones, que de alguna manera reflejan las necesidades de la sociedad al cual está inserida, sobretodo se hablamos en la construcción del conocimiento social.

Los estudios de laboratorio y fabricación del conocimiento científico nos demuestran que no sólo la ciencia con sus innovaciones tecnológicas impactan en la vida cotidiana, como en las dinámicas sociales y en las políticas públicas, por ejemplo (Knorr-Cetina, 1983). El opuesto también es verdadero: los recursos destinados al emprendimiento de una investigación, los esfuerzos ejercidos en prol de temáticas y disciplinas a ser investigadas, el interés de generaciones de científicos, las materias primas y artefactos disponibles para tanto, también son un reflejo de los deseos o necesidades de una sociedad.

El esfuerzo de un científico social que estudia aspectos de la ciencia, debería ser en función de identificar las múltiples afectaciones entre sociedad y laboratorio a la hora de investigar. Un laboratorio no expresa necesariamente la verdad o el correcto funcionamiento de las cosas en su estado natural. Expresan marcas evidentes de intereses consolidados a base de mucho esfuerzo, intencionalidad y construcciones de lógicas que se pasan por racionales y controladas.

Pero sabemos, ya fuimos enseñados, de que el proceso de constitución de los conocimientos, originados en los más distintos laboratorios alrededor del mundo, no son hechos de un caminar completamente racional y en línea recta. Los laboratorios representan un entorno crucial para la generación de conocimiento científico y de innovación, pero también funcionan como espacios donde, a través de la observación atenta e implicada, es posible percibir el proceso de manufacturación del conocimiento mismo, duro y crudo, tal como se presenta frente a los múltiples desafíos del cotidiano y la rigurosidad de las estructuras burocráticas y protocolar que puede volver a ser la vida de un científico.

Muchos de los autores anteriormente citados concordarían con Knorr-Cetina cuando dice que en el interior de los laboratorios parece existir un principio que gobierna la acción del

laboratorio: “preocupación en hacer que las cosas funcionen”, lo cual apuntaría a una predisposición al éxito, más que a la verdad (Knorr-Cetina, 2005).

El vocabulario de los científicos sobre como las cosas funcionan o no funcionan, o sobre los pasos que se dan para hacerlas funcionar, no refleja ninguna forma de verificacionismo ingenuo: se trata, en realidad, de un discurso apropiado a la *manufactura instrumental* del conocimiento en el *taller* llamado "laboratorio". Lograr hacer que las cosas "funcionen" es una búsqueda mucho más mundana que la de la verdad, y una búsqueda en la cual la vida cotidiana de la ciencia es constantemente convertida en créditos mediante la publicación. En consecuencia, es el logro de hacer que las cosas "funcionen" lo que se refuerza como objetivo concreto y factible de la acción científica, y no el distante ideal de la verdad que nunca se aleja del todo. (Knorr-Cetina, 2005, p.59).

Por esta razón entiendo la necesidad de conducir una *etnografía de laboratorio* (Latour y Woolgar, 1979) ya que me fue brindada una posibilidad de estar insertada en unos de estos laboratorios. Entender las dinámicas, las negociaciones, las condiciones de determinación y algunos pormenores, de la producción del conocimiento de mi interés, podría, entonces, ser un objetivo logrado, una vez en contacto con un laboratorio de investigaciones aplicadas a ciencias de datos, multidisciplinar, académico y con vocación a los grandes datos públicos.

Y aunque sea evidente la diferencia de temática, contexto y tipo de laboratorio investigado en este presente estudio, con cualquier otro clásico estudio de laboratorio emprendido en las ciencias sociales, sostengo que estas diferencias no desvalorizan ni comprometen la adecuación de este marco para contextualizar mi investigación.

En similitud a los célebres autores y estudiosos sociales de los estudios de ciencia, tecnología y sociedad, mi intención es observar las prácticas cotidianas, las interacciones sociales y los métodos de trabajo de los científicos en un laboratorio, con el objetivo de comprender cómo el conocimiento es producido, legitimado y, posteriormente, transferido a diferentes esferas de la sociedad. Este análisis de las dinámicas internas y de las interacciones que resultan en la difusión del conocimiento científico, incluso en un campo distinto, refleja el cuidadoso trabajo que todos ellos realizaron en sus estudios de laboratorios. Mi objetivo es captar la fabricación del conocimiento científico no es solo un descubrimiento de hechos, sino una construcción social, material y técnica.

El estudio etnográfico conducido no necesariamente me hizo entender los caminos y arranjos seleccionados detrás de cada uno de los artículos analizados en el estudio bibliométrico. Tal faceta quedó más evidente por medio de las entrevistas realizadas con muchos de los autores de las publicaciones levantadas a partir de la etapa bibliométrica de la investigación en las librerías Scielo y Scopus. Pero sí, permitió una inmersión más profunda en los quehaceres, dinámicas de trabajo y operacionalización de las actividades en un ambiente de investigación en datos masivos y procesamiento computacional de datos interesados en la temática big data.

El laboratorio estudiado en cuestión, mantiene un alto índice de productividad con respecto a otro formato de texto. Los informes de acciones prácticas, o las guías de utilización de datos o de proyectos en ciencias de datos; los relatorios para proyectos más ejecutivos, como Fondef<sup>60</sup> en Chile, por dar un ejemplo, son el tipo de material de difusión y reporte de los resultados de sus investigaciones. Materiales-fuentes analizados como documentos y puestos en perspectiva a la vivencia cotidiana de estos laboratorios y sus grupos de trabajo.

Sin duda, fue la conjunción entre el mapeo de la literatura académica al respecto, más las conversaciones con sus autores, y mi vivencia como investigadora y trabajadora en temas de big data y ciencias de datos, que juntas pudieron producir mi propia mirada un campo disciplinario de las ciencias sociales en plena consolidación.

La presentación de las dimensiones del trabajo investigativo con y a partir de big data, (a partir del capítulo 7) sólo fue posible a través de este exitoso bricolaje, entre técnicas de producción de información y disponibilidad de recursos y contextos. El contacto con el campo trajo la oportunidad de mirar en primera fila los cambios y dinámicas de estructuración de un campo disciplinario naciente y por medio de lo por mi observado, brindarles un capítulo donde me arriesgo a pronosticar hacia dónde se debe encaminar las actividades de nuestro labores, en cuanto científicos sociales en un periodo en la historia del mundo donde la digitalización de la información, su almacenamiento y procesamiento, implica y posibilita otros tipos de habilidades, dinámicas de trabajo y puestos profesionales, ya impactado (para bien y para mal) la manera como producimos ciencias sociales y estudiamos las sociedades a través de ellas.

---

<sup>60</sup> “El Fondo de Fomento al Desarrollo Científico y Tecnológico, Fondef, fue creado en 1991. Su propósito es contribuir al aumento de la competitividad de la economía nacional y al mejoramiento de la calidad de vida de los chilenos, promoviendo la vinculación entre instituciones de investigación, empresas y otras entidades en la realización de proyectos de investigación aplicada y de desarrollo tecnológico de interés para el sector productivo u orientados al interés público”. Información compilada de la página oficial del Gobierno chileno. Accedido en agosto de 2023.

### 6.3 GobLab: un laboratorio social en big data y ciencias de datos

En agosto de 2021, empecé mi trabajo “etnografía de ciencia de datos” en el “GobLab UAI”. El laboratorio estudiado, ya mencionado, es el laboratorio de innovación pública en la Escuela de Gobierno de la Universidad Adolfo Ibáñez en Santiago, Chile. Era lo que decía la descripción en mi boleta de honorarios. Fui contratada para ser parte de un proyecto inédito en Chile, ejecutado por la UAI, con financiamiento de *BID Lab* (el laboratorio de innovación del Grupo BID) y en alianza con socios del mundo público y privado: *Proyecto Algoritmos Éticos, Responsables y Transparentes*. Su objetivo: “instalar capacidades y estándares para incorporar consideraciones éticas en la compra y utilización de inteligencia artificial y algoritmos de decisión automatizada en agencias estatales, y en la formulación y desarrollo de estas soluciones por parte de los proveedores tecnológicos”. (Página web GobLab UAI, accedido en septiembre de 2023)<sup>61</sup>.

Trabajé como “personal de etnografía” por 5 meses. Diseñé el plan de estudio y acompañamiento del proceso de instalación de 2 sistemas algorítmicos diferentes, con propósitos y funcionamientos variados, de dos instituciones públicas distintas: la Defensoría Penal Pública (DPP) y FONASA<sup>62</sup>. Mi objetivo principal era conseguir entender las barreras culturales/sociales a la implementación de estándares éticos en ciencia de datos, y estudiar la interacción entre los funcionarios públicos y los proveedores del sistema de tecnología. Según la dirección del proyecto, en nuestras conversaciones informales se comentaba el esfuerzo realizado para convencer a los demás gestores del proyecto de la necesidad de la contratación de un profesional etnógrafo para la implementación de esta investigación aplicada. Decía que así lo estaban haciendo en los más renombrados laboratorios en ciencias de datos en Estados Unidos, con quienes tenía contacto estrecho y había tenido la oportunidad de conocer en persona. También expresó la urgencia de saber si entre el equipo de desarrolladores del sistema y los usuarios finales de la plataforma conseguirían realmente incorporar la herramienta creada en su trabajo diario.

<sup>61</sup> <https://goblab.uai.cl/algoritmos-eticos/> - video *teaser* del proyecto:

▶ Proyecto Algoritmos Éticos, Responsables y Transparentes

<sup>62</sup> El proceso licitatorio de ambas empresas proveedoras del servicio, bien como las cláusulas del contrato y las especificaciones de las actividades y aportes del proyecto en cuestión son públicas, hacen parte de la ley de transparencia y compras de servicios a entidades públicas y pude ser consultada a través del siguiente enlace: <https://www.iadb.org/es/project/CH-T1246>

Por otro lado, el equipo técnico, a través de los investigadores del laboratorio, asesoraba a los funcionarios públicos que estaban a cargo de los proyectos de ciencia de datos. También tenía como función apoyar la implementación de los estándares éticos que se habían solicitado a los proveedores en las licitaciones de los proyectos.

La idea común era lograr crear circunstancias y procedimientos capaces de consolidar la implementación de sistemas algoritmos éticos y responsables al interior de los servicios, aplicaciones y sistemas a ser desarrollados en función de las necesidades reales de diversos servicios, ministerios, organizaciones o instituciones públicas chilenas. Un proyecto pionero en América Latina y el Caribe, por el cual tengo además, orgullo en haber sido parte.

Uno de los sistemas licitados tenía como meta mejorar y expandir la ejecución de un previo y exitoso proyecto piloto implementado inicialmente en una comuna de Santiago, y luego en varias regiones del país. El sistema a ser desarrollado sería un asistente virtual que auxiliaría a los defensores públicos en la gestión de los procesos y causas bajo la responsabilidad de cada defensor del sistema público de justicia de Chile. La defensoría también pasaba por un proceso de digitalización de los documentos, de las fichas y pruebas de los casos y los imputados. Necesitaban un sistema que alojara apropiadamente tal información (sensible y privada) y que además fuera capaz de prever causas favorables en base a una serie de criterios legales, modelos matemáticos, procedimientos burocráticos y jurídicos<sup>63</sup>, que no ven al caso relatar con mayor detalle, en este momento.

La ejecución del proyecto de la Defensoría Penal Pública, asesorado por el proyecto “Algoritmos Éticos”, estaba basada en la promesa de fabricar un sistema que implementara un algoritmo apto para generar información que auxiliase a los defensores públicos al momento de tomar decisiones con respecto qué “imputados” priorizar en atención, en base a, por ejemplo, gravedad o urgencia del caso, tiempo de espera o la probabilidad de causa favorable. Esto, gracias a la detección de patrones de casos antiguos de la DPP (base de datos proporcionada por la institución), en combinación con la información disponible en la plataforma, del caso y el individuo en cuestión. Se había propuesto, en base a experiencias anteriores, un semáforo de colores<sup>64</sup> que determinara la probabilidad del imputado de obtener la mejor sentencia posible, conocidas como las causas favorables.

---

<sup>63</sup> Para más informaciones consultar las bases del concurso: <https://www.iadb.org/es/project/CH-T1246>

<sup>64</sup>Color verde: positivo a la causa favorable; Color rojo: negativo, o alta improbabilidad de causa favorable y color amarillo: medio del camino, o sea, había trabajo que hace para que la causa fuera favorable)

El segundo proyecto buscaba implementar un sistema de detección preciso que anticipara la probabilidad de fraude en la emisión de los Bono PAD - MLE<sup>65</sup> por parte de los prestadores de servicios médicos a FONASA, el seguro público de salud. Se intentó disminuir la tasa de malversación de los recursos de la institución, el cual se había acrecentado en los últimos años. A su vez, se debía apoyar el trabajo del equipo humano de fiscalización<sup>66</sup>.

Estas licitaciones no fueron adjudicadas, por lo que finalmente fui contratada como investigadora cualitativa en el proyecto Fondef IDEA I+D - SAMSARA<sup>67</sup>, Sistema de Alerta y Monitoreo Satelital de Áreas de Relevancia Ambiental, el cual pretende identificar cambios en la estructura de la vegetación en humedales urbanos, turberas de Chiloé y el bosque y matorral esclerófilo de la región Metropolitana para ayudar a su preservación.

Acá, fui responsable de conducir una encuesta nacional con los servidores y funcionarios públicos involucrados en la utilización de imágenes satelitales: “Encuesta uso de imágenes satelitales en el estado de Chile”, un proyecto liderado por investigadores de la Facultad de Ingeniería y Ciencias y la Escuela de Gobierno de la Universidad Adolfo Ibáñez (UAI), desarrollado en conjunto con la Superintendencia del Medio Ambiente, el Servicio de Evaluación Ambiental y el Ministerio del Medio Ambiente gracias a la adjudicación de un proyecto Fondef.

**Figura 36.** Descripción y colaboraciones proyecto SAMSARA



*Nota.* Captura de pantalla del sitio web de SAMSARA.

<sup>65</sup> Pago Asociado a Diagnóstico, Modalidad Libre Elección. El PAD es un beneficio para usuarios de FONASA. Les permite acceder a paquetes de prestaciones y atenciones en salud pagando un precio fijo. Este beneficio se puede realizar en establecimientos privados de Salud en convenio con FONASA, lo que se conoce como Modalidad Libre Elección (MLE).

<sup>66</sup> Ver licitación ID 591-41-LE21 en [www.mercadopublico.cl](http://www.mercadopublico.cl)

<sup>67</sup> <https://goblab.uai.cl/proyecto-samsara/>

Junto a una colega, éramos la parte cualitativa de este proyecto Fondef. Nuestra misión era entender los principales usos, los desafíos y la transformación que la utilización de la imágenes satelitales proporcionaban al ejercicio de sus actividades. La idea era comprender bien las ventajas y dificultades en utilizar este tipo de tecnología para lograr, en la segunda parte del proyecto, incorporar demandas y casuísticas pertinentes al trabajo real vivido en la vida laboral de estas personas. El principal: desarrollar un algoritmo que funcionara. Eso, en su momento significó, que fuera capaz de procesar la mayor cantidad posible de imágenes satelitales disponibles del territorio chileno en un sólo ambiente, respetando las especificidades y diversidades de tipos y propósitos de uso de estas imágenes. Desafío no menor. O sea, un sistema que llegara más cercanamente a las necesidades de nuestros encuestados.

Después del mapeo de estos usuarios, de la aplicación, del análisis de la encuesta, y de las decenas de entrevistas hechas a estos servidores y sus equipos, confeccionamos un informe que reportaba los resultados, además de lograr establecer los principales “Stakeholders”, parafraseando una vez más vocabulario nativo del laboratorio, o sea: funcionarios y servidores del Estado de Chile- usuarios de imágenes satelitales, lo que fue clave para el establecimiento de un mapa de actores claves, trabajadores de imágenes satelitales, que hasta el día de hoy, siguen movilizados en garantizar la implementación de este sistema, previsto para enero de 2024.

Uno de los principales aportes del GobLab a esta investigación, como consecuencia colateral del estudio cualitativo que emprendimos, fue la capacidad de capitalizar y difundir la conversación con respecto a las imágenes satelitales para y en Chile, fortaleciendo, por otro lado, la construcción de vínculos y redes que viabilizan la mantención de este sistema una vez puesto en funcionamiento por los investigadores de procesamiento de imagen satelital y teniendo como base de partida un sistema australiano.<sup>68</sup>

Estas dos experiencias me llevaron a ser convidada, a fines de 2022 y al término de mis actividades en el proyecto Fondef, a ser miembro integrante del Comité Editorial de otra iniciativa vinculada al GobLab: el *Repositorio Algoritmos Públicos - GobLab - UAI*, que recibió el Premio latinoamericano de democracia digital en 2023. El repositorio describe como:

---

<sup>68</sup>Para saber más y acceder a los resultados de este proyecto:  
[https://goblab.uai.cl/wp-content/uploads/2022/12/Estudio-Uso-Imagenes-Satelitales-Estado-de-Chile\\_UAI\\_final\\_dnov2022.pdf](https://goblab.uai.cl/wp-content/uploads/2022/12/Estudio-Uso-Imagenes-Satelitales-Estado-de-Chile_UAI_final_dnov2022.pdf)

Es una plataforma digital donde se recopila y muestra, de manera centralizada, la información sobre sistemas de soporte o toma de decisiones automatizadas que han sido realizados y/o están siendo utilizados por instituciones públicas en Chile. Nuestro fin es otorgar visibilidad a los sistemas de soporte o toma de decisiones automatizadas en el sector público chileno, buscando incentivar a las instituciones públicas a innovar. Entendemos que el país quiere y debe avanzar en el mejoramiento de la gobernanza y la apertura de datos, por lo que, a través del repositorio, queremos contribuir con tal labor. (Página web Repositorio, accedido en septiembre de 2023).<sup>69</sup>

El comité se reúne una vez al mes para, en base a una metodología y protocolo diseñado por el propio repositorio, decidir en consenso cuáles de los algoritmos previamente levantados por un asistente de investigación, son incluidos o no al acervo del repositorio. Los algoritmos detectados son categorizados y evaluados en términos de concepto, tipos de uso, grado de implementación y automatización del proceso, entre otros.

Mi experiencia investigativa y laboral en el laboratorio fue realizada durante agosto de 2021 a diciembre de 2022, totalizando un periodo de 1 año y 4 meses de investigación y contrato de trabajo. Mientras ejercía mis actividades en los proyectos contratados, mantenía un diario de campo digital, vinculado a mi correo personal donde iba anotando pequeñas notas en un documento de Google Drive instalado en todos mi equipos. Al final de los días de la semana, realizaba un resumen de las actividades hechas, como forma de mantener actualizado este documento doctoral, y también como documento de seguimiento de mis actividades laborales.

En muchas ocasiones, durante el tiempo que estuve trabajando en el proyecto *GobLab BID\_UAI*, los demás compañeros no entendían muy bien cual era mi rol ni lo que se suponía que debería estar haciendo. Escribir los principales puntos y actividades realizadas día a día, fue una de las estrategias que implementé para comprobar y justificar la importancia de mi participación en el equipo y el valor de conducir un estudio de esta naturaleza en tal contexto. Me alegro de haber mantenido este registro. Hoy son parte de una documentación precisa de mi jornada profesional y académica.

Pero no es solamente por mi trayectoria laboral que se justifica la elección del GobLab UAI como mi caso de estudio. Este laboratorio específico, además de todas las actividades de investigación aplicada que llevan a cabo, se constituye como un prestigiado espacio de

---

<sup>69</sup> <http://algoritmospublicos.cl/>

convergencia entre científicos, tecnólogos y agentes de/en/para las políticas públicas, la innovación y la ciencia de datos. Es reconocido en toda América Latina y ganador de premios nacionales e internacionales<sup>70</sup> por el carácter innovador de sus proyectos e investigaciones aplicadas, y por ser unos de los pocos laboratorios en Latinoamérica con enfoque declarado en ciencia de datos para datos de gobierno.

**Figura 37.** *Presentación GobLab, Escuela de Gobierno*

“Nos aliamos con la Escuela de Gobierno UAI porque no hay otra universidad que esté trabajando estos temas en la región de América Latina.”

-- María Fernanda Trigo  
Directora Depto. Gestión Pública Efectiva de la OEA – 22 agosto 2022

*Nota.* De la presentación de final de año del Laboratorio

Su relevancia en el escenario nacional chileno y en la región puede comprobarse por su amplia red de colaboradores y organizaciones de las que es parte. GobLab hace parte de tres redes de innovación pública: La Red Innolabs, red iberoamericana de innovación pública impulsada por Nova GobLab de España y financiada por el Programa Iberoamericano de Ciencia y Tecnología para el desarrollo, CYTED.

También de la Red UnInPública (Red de Universidades por la Innovación Pública) que tiene como misión dar respuesta a la demanda social de transformación a través del conocimiento. Y la Red Iberoamericana de Laboratorios impulsada por el proyecto Innovación Ciudadana de la Secretaría General Iberoamericana de la Secretaría General Iberoamericana.

**Figura 38.** *Redes internacionales asociadas a GobLab*



*Nota.* Imágenes extraídas del sitio web de GobLab.

<sup>70</sup>A nivel nacional obtuvo el [premio “Academia” por su apoyo a niñez y juventud LGBTIQ+](#) por parte de la Fundación Todo Mejora y su director, María Paz Hermosilla, se incorporó al Centro Nacional de Inteligencia Artificial como colaboradora asociada. A nivel internacional, el Proyecto Algoritmos Éticos fue destacado por el Instituto Nacional de Ciencia de Datos e Inteligencia Artificial del Reino Unido como un [ejemplo de justicia de datos](#). También recibió el [Premio Latinoamericano Democracia Digital](#), y finalmente, la directora del GobLab fue reconocida entre las [100 mujeres líderes a nivel mundial para 2022 de Inteligencia Artificial y Ética por la Women in AI Ethics](#).

Con respecto a la dinámica de mi trabajo, como aún vivíamos las restricciones sanitarias, gran parte de mi actividad fue hecha en ambiente digital en modalidad de teletrabajo. Al incorporarme al laboratorio recibí un email institucional, también acceso a la plataforma de videoconferencia y tuve que incorporarme a una tercera plataforma de gestión de trabajo, el Microsoft Teams. No creo que esto pasaba solamente por cuestiones de comodidad por parte del laboratorio, sino también por el hecho que muchos de sus integrantes se encontraban en diversas partes del mundo, haciendo sus distintas especializaciones, trabajos y propia formación o trabajo.

En todo caso, les comparto algunas impresiones a título de curiosidad: Las instalaciones nuevas de la universidad, el “edificio F” del campus de Peñalolén de la UAI recién se había construido e inaugurado. Todo muy moderno y tecnológico. El espacio destinado a los trabajadores del GobLab estaba en una sala bien iluminada y equipada de computadores, escritorios, armarios y puestos de trabajo. Para entrar libremente al local, había que registrar la huella digital del pulgar. En la derecha de la sala, había otro conjunto de escritorios ocupados por otro personal de la universidad. La sala de la directora del laboratorio se sitúa a parte de los otros, al final del conjunto de oficinas, es amplia y también bien equipada. Ítems de trabajo como computadores, libreros, teléfono, artículos personales de la directora, diplomas, libros, mascota de su equipo de fútbol, un recipiente para almacenar galletas, sobrecitos de té y un hervidor para las once, componen el pasaje de su oficina. Estaba disponible para nuestra utilización impresoras, los propios computadores y las instalaciones de la universidad, de muy buena calidad.

Yo iba presencialmente al laboratorio una vez por semana, todos los miércoles, jornada completa. Era poca la presencia personal de los colaboradores comparados a las reuniones por *Zoom*. Ir al laboratorio me posibilita vivenciar algunas instancias informales de conversación con diversos funcionarios del laboratorio y nada más. Aunque estuvieran 5 personas en la sala, las reuniones seguían en modalidad en línea, aunque la participación física en la sala era reiterativamente mencionada por la directora. “No es obligatorio venir, pero se aprecia mucho”, decía la Directora del GobLab, tratando de motivar a la gente a comparecer.

Es por esto la notoria y más contundente presencia virtual y trabajo online que presentaré en campo a partir de catectización y descripción de la página web del laboratorio, una valioso

documento que evidencia la narrativas y percepciones que el laboratorio tiene y por el que se hace conocer.

El sitio web del GobLab es un soporte internamente valorado. Teníamos reuniones semanales de coordinación y muy periódicamente discutimos sobre la usabilidad y eficiencia de la página, mejoras que podrían ser hechas y la búsqueda de estrategias para incorporar a alumnos de pasantías expertos en comunicación a modo de minimizar los costos de manutención de la página, de la manera más óptima posible. El laboratorio cuenta con un practicante dedicado a la manutención de ella, además de un presupuesto destinado a contratación de servicios de hospedaje del dominio de la página, marketing digital y difusión vía redes sociales de sus actividades.

### ***6.3.1 Uso de datos y la innovación pública: como se construye un laboratorio de ciencia de datos de gobierno***

El laboratorio GobLab tuvo sus inicios influenciado por el deseo del, en su momento, decano de la Escuela de Gobierno de la UAI, por implementar un centro que apoyara la modernización del Estado. Después de un proceso de un año y medio de conversaciones y negociaciones entre él y la fundadora, actual directora del laboratorio, el GobLab fue oficialmente creado en el año 2017. Su historia comenzó, por lo tanto, a partir de la visión de promover la innovación pública y el uso de datos en las políticas gubernamentales.

María Paz Hermosilla, fundadora del laboratorio, trajo desde su experiencia y trayectoria en la administración pública, la misión y el interés por la transformación del Estado. Durante su tiempo en la administración pública, desarrolló una pasión por la modernización del Estado, los datos abiertos, la transparencia y el uso responsable de los datos en la gestión pública. En búsqueda de conocimientos adicionales, obtuvo una maestría en administración pública en la NYU (New York University Public Policy), donde también empieza a trabajar como analista en el The Governance Lab (GovLab)<sup>71</sup>, un laboratorio de innovación de la Escuela de Ingeniería en la referida universidad.

Fue en el GovLab de la NYU donde adquirió experiencia en temas relacionados con ciencia de datos y trabajó en proyectos relacionados con *crowdsourcing*, innovación pública y *smart cities*. Mientras estudiaba y trabajaba en GovLab, María Paz se cruzó con Ignacio Briones,

---

<sup>71</sup> <https://thegovlab.org/> - The Governance Lab

entonces decano de la escuela de gobierno UAI, quien tenía la visión de crear un centro para la modernización del Estado en la Escuela de Gobierno de la Universidad Adolfo Ibáñez.

Juntos comenzaron a explorar la idea de establecer un laboratorio de innovación en Chile, y después de una serie de acomodamientos institucionales, en junio de 2016, María Paz acepta la invitación de la Escuela, y a finales de ese año se traslada a Chile fundando el GobLab en la Escuela de Gobierno de dicha Universidad.

Álvaro Bellolio, quien venía volviendo de la Universidad de Chicago, y había dedicado su MPP a temas de data science. Al unirse, decidieron centrarse en la ciencia de datos como la metodología principal del laboratorio. Así nació la unión entre ciencia de datos, gestión pública y ética de datos, enfoque y línea de trabajo declarada del laboratorio<sup>72</sup>.

El laboratorio GobLab fue fundado con el objetivo de autofinanciarse mediante programas de educación, proyectos y consultorías. Inicialmente, el presupuesto era limitado, pero con el tiempo fue creciendo a medida que nuevos proyectos se incorporaron a sus actividades. Hoy día mantienen una operación capaz de responder a los gastos operativos y de personal de laboratorio.

### ***6.3.2 Formar, incidir y investigar: la tríade que sostén la misión del laboratorio***

Durante mi proceso de investigación en GobLab, tuve acceso a diversos documentos internos del laboratorio. Participé de las reuniones semanales de coordinación, a las que se sumaban algunas actividades relacionadas con la difusión y producción de materiales internos al laboratorio. En una ocasión me tocó acceder al documento “Memoria Anual 2022”, un informe de las actividades realizadas por el laboratorio en el año 2022. La siguiente cita, corresponde a uno de sus apartados, donde puede leerse sobre su motivación. Por otra parte, el GobLab despliega su quehacer en 3 áreas, con sus respectivos objetivos estratégicos que detallo en la tabla 21.

“La misión del GobLab UAI, laboratorio de innovación pública de la Escuela de Gobierno, es contribuir a la innovación en políticas públicas para beneficiar a la sociedad. Su visión a 5 años, declarada en el 2020, es *ser reconocidos por contribuir*

---

<sup>72</sup> “En esta línea de trabajo exploramos las implicaciones éticas y sociales del uso de la ciencia de datos, sus efectos en la privacidad, transparencia, equidad y derechos humanos, además de explorar las herramientas para hacer un uso ético y responsable de los datos” (Pagina web GobLab)

*a la transformación del sector público y al impacto social mediante ciencia de datos utilizada de manera ética y responsable.”*

**Tabla 21.** *Objetivos estratégicos GobLab, 2023*

Área	Objetivo estratégico
Formación	Proveer capacitación sustantiva y aplicada para servidores públicos de América Latina para la innovación pública, con un foco en programas articulados desde cursos y diplomados a magíster.
Investigación aplicada	Mejorar la gestión del Estado y las políticas públicas mediante el uso de evidencia aportada por investigación interdisciplinaria, aplicada y asociativa.
Incidencia	Impulsar la creación de una gobernanza integrada de datos en Chile que contribuya al desarrollo ético de sistemas de tomas de decisión, mediante el trabajo colaborativo y compartiendo buenas prácticas con la región.

*Nota.* Resumen de la memoria 2022 de GobLab UAI.

GobLab, entonces, tiene tres pilares: formación, investigación aplicada e incidencia. Estos tres fundamentos se sostienen entre sí, y según mi parecer y vivencia en el laboratorio, se complementan de manera simbiótica, sin que uno sobreviva sin el otro. El enfoque en la formación y capacitación de profesionales y servidores públicos en temáticas de innovación y ciencia de datos, no es solamente responsable por mantener financieramente las actividades del laboratorio a través de los pagos y costos de los cursos, magister y diplomados ofrecidos, sino también de donde nacen muchas otras actividades.

De los cursos de formación es que provienen la mayor parte de los proyectos en investigación aplicada que el laboratorio desarrolla. Los contratos de colaboración y ejecución de innumerables proyectos son provenientes de ex-alumnos de sus diplomados y cursos. No es común escuchar en las reuniones que un proyecto tal haya sido aprobado y puesto en marcha gracias a uno de sus alumnos, que durante sus estudios, percibieron alguna necesidad inmediata en su institución. Son ellos los que internamente en sus servicios, son capaces de movilizar la organización con el objetivo de poner en actividad los diseños de los proyectos que engendrado en su paso por el laboratorio.

Una vez establecido el acuerdo de colaboración, GobLab acababa asociándose con uno de los colaboradores. Por lo general prestando consultorías y apoyo en temas de ciencias de datos, o incorporando en la ejecución de los proyectos protocolos de implementación del procesamiento de datos con base en la ética, privacidad y responsabilidad con los datos

(estudio de sesgo) en los modelos y sistemas a ser implementados. Es de esta manera que el laboratorio logra difundir conocimiento, ejecutar proyectos e incidir de igual manera y concomitantemente en el sector público chileno, haciendo cumplir su misión de existencia.

### 6.3.3 Colaboradores diversificados como fortaleza institucional

Durante el 2022, el año que estuve presente con más frecuencia en el laboratorio, el equipo del GobLab estuvo compuesto por un grupo diverso de colaboradores en el desarrollo y la ejecución de los proyectos del laboratorio. A continuación, se presenta la composición del equipo, sus respectivos cargos y formación:

**Figura 39.** *Equipo GobLab, 2022*

Como se puede constatar, 14 personas componían el cuadro de funcionarios del laboratorio, en 2022. Con respecto a los investigadores asociados de la Universidad

1. **Directora** - Máster Administración Pública
2. **Subdirectora** (Desde noviembre de 2022) - Abogada / Máster Derecho de las Nuevas tecnologías
3. **Coordinadora de Gestión** - Administración Pública y Licenciada en Ciencias Políticas y administrativas
4. **Coordinadora de Formación** - Ingeniero Comercial - MBA
5. **Asesor I+D** - Bachiller y Magíster en Ciencias de la Computación
6. **Analista** - Sociólogo
7. **Jefa de proyecto** Algoritmos Éticos - Ingeniera Civil - MBA
8. **Investigador Postdoctoral** proyecto Algoritmos Éticos - PhD en Ingeniería línea informática y Magister en Ingeniería Automatización Industrial.
9. **Etnógrafo e Investigador Cualitativo** proyecto Algoritmos Éticos - Sociólogo
10. **Encargada de comunicaciones** proyecto Algoritmos Éticos - Periodista
11. **Investigadora Proyecto Samsara** - Magíster en Sociología y candidato a doctor en Ciencias Sociales
12. **Jefa de proyecto** Reducción de la Malnutrición Infantil en Chile - Ingeniera Informática
13. **Consultor asociado proyecto CAF-Subdere** - Ingeniero Civil en Computación, U. Chile.
14. **Asistente de comunicaciones** - Periodista

(UAI), se incorporan como investigadores asociados Naim Bro, profesor asistente de la Escuela de Gobierno, e Isidora Paiva, profesora instructora de la Escuela de Psicología, entre otros personajes prestigiosos de la institución. Con eso el GobLab alcanzó 12 investigadores/as asociados/as, como puede verse en la siguiente figura.

**Figura 40.** Investigadores asociados a UAI en GobLab en 2022

1. Luis Herskovic, Profesor Asistente, Escuela de Gobierno
2. Luis Santana, Profesor Asistente, Escuela de Comunicaciones y Periodismo
3. Mayra Fedderson, Profesora Asistente, Facultad de Derecho
4. Florencia Borrescio, Profesora Asociada, Escuela de Negocios
5. Nieves Valdés, Profesora Asociada, Escuela de Negocios
6. María Josefina Escobar, Profesor Asistente, Escuela de Psicología
7. Matías Garretón, Profesor Asociado, Escuela de Diseño
8. Daniel Loewe, Profesor Titular, Facultad de Artes Liberales
9. Naim Bro, profesor asistente de la Escuela de Gobierno
10. Isidora Paiva, Profesora instructora de la Escuela de Psicología
11. Gabriela Denis, Master en Economía Aplicada UCLA, Estudiante Doctorado U. Chile
12. Claudio Aracena, Máster en Tecnologías de la Información y Máster en Administración en Tecnologías de la Información.

**Figura 41.** Colaboradores UAI en GobLab en 2022

1. Cristián Rettig, Profesor Asistente, Facultad de Artes Liberales
2. Arturo Arriagada, Profesor Asociado, Escuela de Comunicaciones y Periodismo
3. Gonzalo Ruz, Profesor Titular, Facultad de Ingeniería y Ciencias
4. Sebastián Moreno, Profesor Asociado, Facultad de Ingeniería y Ciencias
5. Miguel Carrasco, Profesor Asociado, Facultad de Ingeniería y Ciencias
6. Mauricio Galleguillos, Profesor Asociado, Facultad de Ingeniería y Ciencias
7. Javier Lopatin, Profesor Asistente, Facultad de Ingeniería y Ciencias
8. Fernanda Díaz, Profesora Asistente, Escuela de Psicología
9. Alejandra Abufhele, Profesor Asistente, Escuela de Gobierno

*Nota.* El GobLab realizó proyectos de investigación con estos 9 profesores de UAI durante 2022.

Finalmente, los 9 ayudantes, 5 de ellos estudiantes de UAI y uno parte del programa MISTI (International Science and Technology Initiatives )<sup>73</sup> del MIT (Massachusetts Institute of Technology); 2 de la Universidad de Chile y 1 de la Pontificia Universidad Católica de Chile.

A través de las tablas y la presentación de funcionarios, asistentes, colaboradores y asociados, podemos ver, por un lado, la diversidad de disciplinas presentes en la formación de sus colaboradores: El GobLab cuenta con un equipo multidisciplinario de miembros que tienen antecedentes académicos en áreas que van desde la administración pública, derecho, artes liberales, hasta la ciencia de la computación, ingeniería, la sociología, psicología, diseño y las

<sup>73</sup> <https://misti.mit.edu/>

ciencias sociales. Esta diversidad de disciplinas enriquece el enfoque del laboratorio y su capacidad para abordar diversos aspectos de la innovación pública.

El segundo factor a ser observado es la especialización de sus colaboradores a temas relevantes y vinculados al laboratorio. El personal del GobLab demuestra una especialización en temas importantes a la innovación pública, como, por ejemplo la ética de los algoritmos, la automatización industrial de procesos, tecnología de la información, negocios aplicados, la sociología y la comunicación. Esta formación diversa y al mismo tiempo especializada (científicamente y proveniente de los patrones académicos) contribuye con la experiencia necesaria para abordar proyectos y desafíos complejos en el ámbito gubernamental. Vinculado a esta dimensión, podemos apreciar el alto nivel académico de sus colaboradores: La mayoría de ellos posee títulos de posgrado, como maestrías y doctorados, lo que evidencia el grado de exigencia en términos de formación para su incorporación a los proyectos del laboratorio. Una segunda explicación posible a este requisito, es la necesidad de profesionales especializados y en constante formación que demanda este campo del saber.

Otro factor que es posible notar es el claro enfoque en temáticas de datos y tecnología. Se observa que varios miembros del equipo tienen antecedentes académicos en áreas o temáticas afines a tecnología, informática y ciencia de datos. Yo misma pasé por el mismo criterio de selección. Un colega me contó, durante una conversación en la hora del café, que uno de los factores que más contribuyó para mi contratación fue el hecho de estar cursando un proyecto de tesis en big data y ciencias sociales. Al parecer, era la única candidata vinculada a temáticas de tecnología, más allá de los estudios de redes y medios sociales.

Además, puede percibirse una complementariedad de roles en la formación de sus colaboradores, reflejando su capacidad para abordar una amplia gama de desafíos relacionados con el uso de datos en el gobierno. Esta combinación de conocimientos y experiencia es un recurso importante para el laboratorio en su búsqueda de promover la innovación en el sector público.

Es por esto que, en conjunto, el alto nivel de formación académica de los colaboradores del GobLab y la combinación de diversos conocimientos y habilidades académicas presentada por sus colaboradores, que el laboratorio se ha convertido en una entidad capacitada para conducir investigaciones de alta calidad, a la vez que desarrolla proyectos innovadores e incidentes.

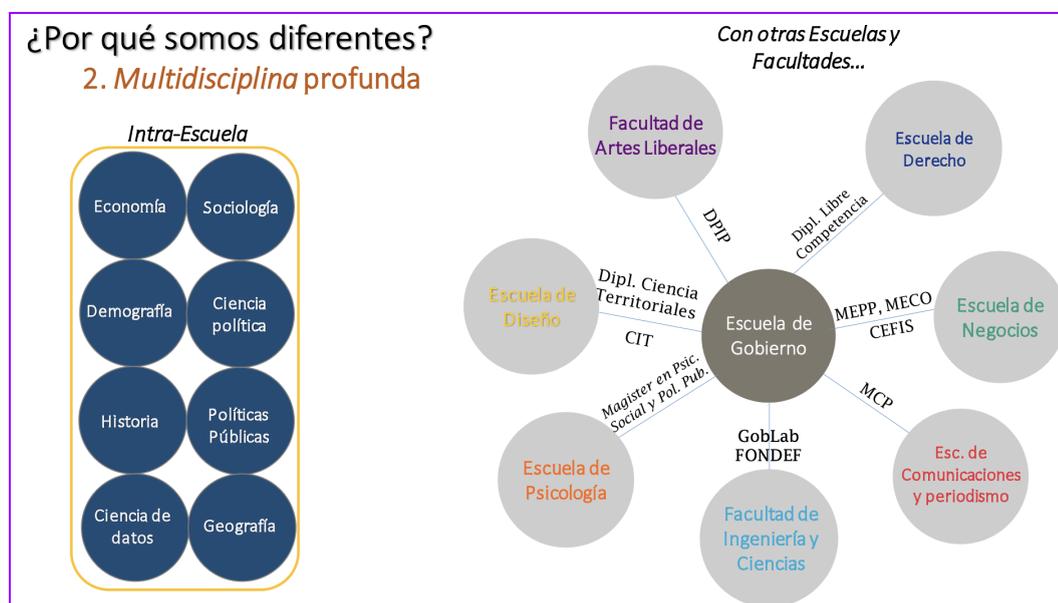
**Figura 42.** *Colaboradores GobLab en 2022*



*Nota.* Fotografía de archivo interno de GobLab

Esta historia no podía ser diferente, siendo el GobLab vinculada a la Escuela de Gobierno de la UAI, una de las facultades más multidisciplinarias de esta Universidad. En una de las reuniones semanales de coordinación del laboratorio, nos tocó acompañar la presentación de la estructura de la Escuela de Gobierno, por parte de la decana de la Escuela, donde mostraba las diferencias con relación a otras facultades semejantes en América Latina. En una de las láminas (figura abajo), encontramos la presentación de las demás escuelas y facultades comparadas.

**Figura 42.** *Centros colaboradores Escuela de Gobierno UAI, 2022*



*Nota.* Presentación Escuela de Gobierno UAI a GobLab

Esta imagen ilustra la profunda multidisciplinariedad que tanto la Escuela de Gobierno como GobLab poseen, justificando desde su origen la vocación con relación a la interdisciplinariedad de sus actividades, formaciones y proyectos ejecutados.

#### **6.3.4 La formación y capacitación humana como práctica de consolidación de un saber**

Como vimos, desde 2016 el proceso de formación es un pilar importante en GobLab y, en este sentido, ha desarrollado una sólida estrategia que ha evolucionado con el tiempo. En un principio, se enfocaron en la creación de diplomados y cursos para capacitar a profesionales en el sector público. En la actualidad, ofrecen varios programas de formación que abordan distintos aspectos de la administración pública y el uso de datos en el gobierno.

Desde su inicio hasta la actualidad GobLab ha promocionado decenas de cursos, entre ellos, los mencionados en la figura abajo.

**Figura 43.** *Formación GobLab UAI 2022*

<p>Curso Automatización de Reportes con R;  Curso Gobernanza de Datos;  Curso Evaluación de Proyectos de Ciencia de Datos;  Ley de transformación digital;  Curso Introducción a R para Análisis de Datos;  Diplomado en Ciencia de Datos para Políticas Públicas;  Curso Inspecciones Proactivas con Ciencia de Datos;  Curso Formulación de Proyectos de Ciencia de Datos;  Introducción a Python: con enfoque para las políticas públicas;  Ciencia de Datos para Directivos Públicos</p>
--

El laboratorio comenzó ofreciendo un diplomado (Diplomado en “Big data para las políticas públicas”, actual “Diplomado en Ciencia de Datos para Políticas Públicas”) que se ha mantenido durante siete años y que, en 2024, entrará en su octava versión. Además, han introducido otros dos diplomados, uno orientado a directivos públicos y otro dirigido a analistas.

El enfoque del Laboratorio GobLab no se limita sólo a la capacitación, ya que también han establecido vínculos y colaboraciones con los exalumnos de sus programas. Esto ha llevado a la realización de proyectos de investigación y proyectos aplicados en el sector público. Un ejemplo destacado es un curso diseñado para directivos públicos que se basa en una metodología desarrollada en el Center for Data Science and Public Policy de la Universidad de Chicago. Este curso, adaptado y enseñado en Chile, ha servido para generar vínculos y proyectos con exalumnos y profesionales del sector público. Con el tiempo, GobLab ha ampliado su enfoque hacia la implementación de proyectos, y han creado guías y recursos que abordan todo el ciclo de vida de un proyecto, desde el diseño hasta la ejecución.

En el año 2022, el laboratorio tomó la decisión de transformar sus programas de formación<sup>74</sup>, migrando desde modalidades presenciales e híbridas a un formato completamente en línea y sincrónico, con el objetivo de ampliar el alcance de sus programas y de brindar la oportunidad de participación a estudiantes de regiones de Chile y a personas fuera del país interesadas en su oferta de formación.<sup>75</sup>

Además la red de contactos del Laboratorio GobLab en la administración pública chilena se ha fortalecido gracias a sus ex alumnos, quienes han desempeñado un papel fundamental en la expansión de sus iniciativas de formación y proyectos de investigación. Este enfoque integral en la formación y la colaboración con profesionales del sector público ha contribuido significativamente a la misión del laboratorio de promover la innovación pública y el uso de datos en el gobierno.

De manera paralela, con el propósito de fortalecer los vínculos que han sido fundamentales para la colaboración exitosa con ex estudiantes en proyectos de investigación aplicada, se

---

<sup>74</sup> A lo largo del año 2022, el GobLab ofreció una variada gama de cursos: En mayo, se realizó el curso "Automatización de Reportes con R" en modalidad en línea sincrónica, con una participación de 7 estudiantes; durante agosto, "Formulación de Proyectos de Ciencia de Datos" en formato en línea sincrónico, con un total de 30 matriculados. Este curso era exclusivo para el Servicio Civil. También en agosto, en colaboración con el Tecnológico de Monterrey, se llevó a cabo el curso "Inspecciones Proactivas con Ciencia de Datos" en modalidad en línea sincrónica, con la participación de 10 estudiantes. Durante octubre, se ofrecieron dos cursos en línea sincrónica: "Gobernanza de Activos de Información" con 14 matriculados y "Evaluación de Proyectos de Ciencia de Datos" con 9 estudiantes. En noviembre, se repitió el curso "Automatización de Reportes con R", atrayendo a 13 participantes; y a lo largo de todo el año, estuvo disponible el curso "Introducción a R para Ciencias Sociales" en modalidad en línea asincrónica, con una matrícula de 25 estudiantes. Esto suma un total de 108 profesionales capacitados por el Laboratorio.

<sup>75</sup> El 2022 se realizó la sexta versión del Diplomado en Ciencia de Datos para Políticas Públicas, con 21 matriculados provenientes de instituciones como el Servicio Nacional de Migraciones, INE, SUSESO, Centro Nacional de Conservación y Restauración, CMF, JDRF UK, JUNAEB, Superintendencia de Educación Superior (SES), SENAME, CORFO, Servicio Civil, Fiscalía Nacional Económica y la Secretaría de Hacienda (Colombia), entre otras. En agosto de 2022 comenzó la primera versión del Diplomado Internacional en Ciencia de Datos para la Dirección Pública, en alianza con la Escuela de Gobierno del TEC de Monterrey, con 19 matriculados (Info Memorial 2022, documentos internos de GobLab).

implementaron dos jornadas de vinculación presenciales al año. Estas jornadas se diseñaron exclusivamente para estudiantes y ex estudiantes del GobLab, consolidando así la creación de la Comunidad de Innovación Pública del GobLab UAI.

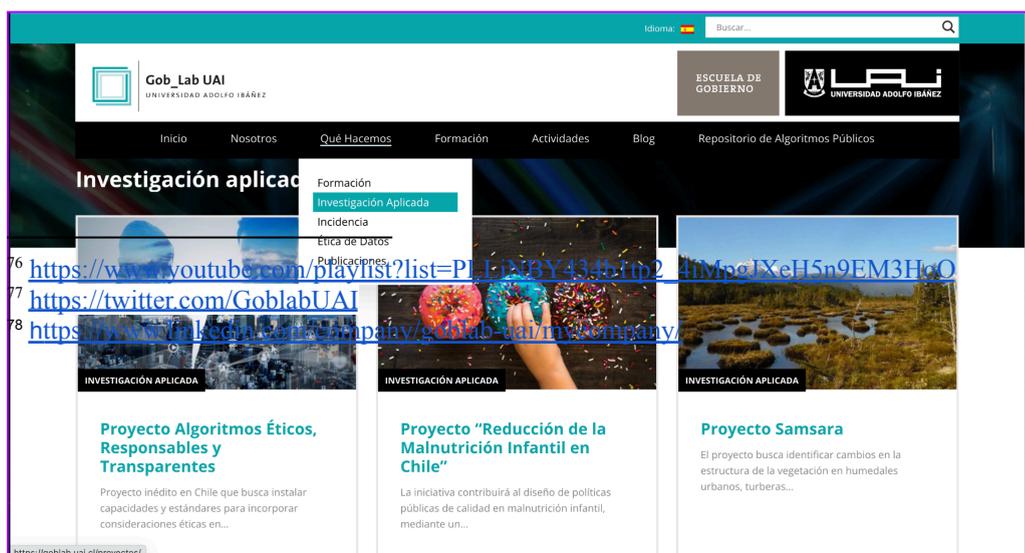
Un último aspecto que me gustaría mencionar es con relación a un segundo canal de formación que hace parte del laboratorio: aunque no sea reconocido de manera privilegiada en las universidades de toda latinoamérica, GobLab posee una marcada atención a sus redes sociales. Más que ser un canal de comunicación con el público, y futuros estudiantes los canales sociales de GobLab, conforman un verdadero acceso gratuito de formación y difusión del conocimiento.

Manteniendo su canal en Youtube<sup>76</sup>, X<sup>77</sup> y LinkedIn<sup>78</sup> activos, GobLab ha atraído investigadores y colaboradores de varias partes del mundo, al menos bastante en la región. Estos medios funcionan como un canal de acceso gratuito a la formación con numerosas personalidades del mundo de la ciencia de datos, grandes datos de gobierno e innovación en políticas públicas. Hay decenas de horas de formación gratuita provenientes de charlas, seminarios, eventos disponibles en esta plataforma, la cual es hoy una vitrina digital para el trabajo y material que ha realizado por siete años.

### 6.3.5 La Investigación aplicada como principio guía de las prácticas del laboratorio

El segundo pilar de las actividades del laboratorio, son las investigaciones aplicadas. Su página web contiene una sección enteramente dedicada a ella. Después de acceder al menú “Qué Hacemos”, se puede hacer clic en la pestaña “Investigaciones aplicadas”. Una nueva página se abre y 16 proyectos, simétricamente distribuidos a lo largo de la página se presentan con títulos y una breve descripción. Se puede pinchar la imagen correspondiente a cada proyecto para conocer de forma más detallada informaciones sobre cada uno de ellos.

**Figura 44.** Ejemplos de investigación aplicada en GobLab



*Nota.* Captura de pantalla de la página web de GobLab.

La diversidad y cantidad de proyectos de investigación aplicada ejecutados por el laboratorio, fue desde un principio orientada a una diversidad de temas y a su implementación en la vida real de las instituciones públicas, más que en investigaciones teórico-conceptuales. Hay una fuerte ponderación en cuanto a la palabras “aplicada” junto al término “investigación”. Percibí una constante necesidad de demarcar el enfoque práctico y ejecutivo del laboratorio, como misión y vocación. Más que un laboratorio de pensadores, diría que el GobLab es un laboratorio de ejecutores, apasionados por el hacer práctico:

Efectivamente todos tienen que ver con los datos, todos los proyectos de investigación aplicada, que tenemos hasta la fecha, en los últimos 6 años, tenemos 20 proyectos de investigación aplicada y efectivamente toda nuestra investigación aplicada ha sido en temas de datos, principalmente ciencia de datos, pero ahora último, por ejemplo, más de estrategias de cómo usar los datos en general, como que no son de crear algoritmos, eso sería como el macro tema, pero estamos especializados en el uso de datos en el sector público y en esos años. (MPH, Entrevista 32/2022 - Chile. Mi traducción).

**Figura 45.** *Investigaciones GobLab UAI en 2022*

1. Proyecto Algoritmos Éticos, Responsables y Transparentes - Derecho + Protección Social + Salud
2. Proyecto “Reducción de la malnutrición infantil en Chile” - Salud
3. Proyecto Samsara - Preservación Ambiental + Gestión Pública
4. Desarrollo de una Estrategia para el uso de tecnologías Digitales para el fortalecimiento de la Gestión descentralizadas en los gobiernos Regionales - Gestión Pública
5. Diseño de Proyecto de Ciencia de Datos con Perspectiva de Género en el Sector Público - Gestión Pública
6. Convenio de Colaboración UAI - Serpesca - Preservación Ambiental
7. Proyecto Alan Turing - Ética de datos/ temas legales
8. Proyecto Salud Mental Adolescente - Salud
9. Transparencia algorítmica en el sector público - Gestión Pública

10. Convenio de Colaboración UAI - Superintendencia de Salud - Salud
11. Inteligencia Artificial para el Cuidado del Medio Ambiente y Salud de la Población - Preservación Ambiental + Salud
12. Proyecto Crowd Law - Gestión Pública
13. Alerta niñez - Protección Social
14. Modelo predictivo para la prevención del delito (lo barnechea) - Seguridad
15. Costos socioeconómicos de las enfermedades onco-hematológicas pediátricas en Chile - Salud
16. Filantropía de datos - datos privados para fines públicos

Cada uno de estos proyectos acaban por impactar de una u otra forma las actividades del laboratorio. De igual manera, derivan a una serie de productos de distintas naturalezas conforme las exigencias de cada proyecto: guías de recomendaciones, informes con resultados de investigaciones cuantitativas y cualitativas, modelos de predicción éticamente implementados y sistemas operacionales para servidores y servicios públicos son algunos ejemplos de “productos”, para usar la palabra nativa, los cuales son entregados de manera formal adjuntado a cada proyecto; además se suben a internet y difunden para su uso en la vida cotidiana del laboratorio.

El laboratorio también posee un LinkedIn<sup>79</sup> bastante activo donde hacen la difusión de futuros eventos, seminarios, y charlas. También emiten comunicados de contratación de personal, notas de divulgación de acuerdo, alianzas y proyectos firmados, es un medio para difundir y promocionar los materiales, informes, guías y videos didácticos que van siendo publicados al interior de las actividades de cada proyecto.

Me gustaría destacar especialmente algunas capacidades que el laboratorio fue desarrollando como consecuencia del involucramiento de sus investigaciones aplicadas. Primero, mencionar la diversidad de temas que trabajan. Los proyectos de 2022, por ejemplo, abarcan áreas tan distintas entre sí como derecho, salud, preservación ambiental, políticas públicas, seguridad y otras. Esto muestra el compromiso de GobLab por abordar desafíos gubernamentales multifacéticos y contribuir a una variedad de campos.

Por otro lado están las colaboraciones estratégicas con diversas organizaciones, como los convenios de colaboración con Sernapesca y la Superintendencia de Salud. Estas asociaciones amplían la capacidad de GobLab para llevar a cabo investigaciones y proyectos de impacto social e incidencia. Y por fin, el enfoque en la ética y la transparencia como en

<sup>79</sup> <https://www.linkedin.com/company/goblab-uai/>

"Proyecto algoritmos éticos, responsables y transparentes" y "Transparencia algorítmica en el sector público" reflejan la preocupación y consecuente movimiento de promoción de temáticas de este género tanto en el espacio académico- intelectual chileno, como en el sector público.

En resumen, durante 2022, el GobLab UAI estuvo desarrollando 10 proyectos de investigación aplicada en alianza con 19 entidades públicas y con profesores de diversas Escuelas y Facultades: Ingeniería y Ciencias, Comunicaciones, Derecho, Psicología, Negocios y Artes Liberales. Cinco de ellos cuentan con financiamiento internacional, proveniente del BID, CAF, Alan Turing Institute, A + Alliance, Fondo Lacuna y Open Government Partnership. Los proyectos buscan destacar el potencial que tiene la ciencia de datos para resolver problemas públicos, en particular en el campo de la fiscalización, y también abarcan temáticas de ética y gobernanza de datos.

### ***6.3.6 La incidencia como mecanismo de aporte social***

“Buscamos influir en los tomadores de decisiones, promoviendo cambios en políticas públicas relacionadas con datos para el beneficio de la sociedad, interactuando colaborativamente con las instituciones del Estado, otros actores del mundo académico y la sociedad civil” (Página de inicio del sitio web GobLab UAI, 2023).

Este es el primer contenido que uno encuentra al entrar en la pestaña “incidencia” en la página de GobLab. De hecho, queda en evidencia cómo esta dimensión es preponderante e incluso determina la mayoría de las actividades del laboratorio. La impresión es que las actividades de incidencia confieren el tono de funcionamiento general del laboratorio, las contrataciones e inclusión de personal colaborador y financiamiento de las demás actividades.

El quehacer de incidencia de GobLab se divide en tres áreas estratégicas: a) promover el uso de ciencia de datos como herramienta de innovación pública; b) la integración de estándares éticos en el uso de datos en el sector público; c) promover la creación de una Infraestructura de Datos Integrados (IDI) en Chile. Cada una de estas áreas estratégicas, en el año 2022, estuvieron cubiertas con alguna actividad o proyecto de investigación aplicada vigente.

Con respecto a la primera, GobLab llevó a cabo diversas iniciativas en las áreas de divulgación e investigación. Una de las destacadas fue el lanzamiento del "Repositorio Algoritmos Públicos"<sup>80</sup>, una plataforma digital que centraliza información sobre sistemas de

---

<sup>80</sup> [Repositorio Algoritmos Públicos](#)

toma de decisiones automatizadas utilizada por instituciones públicas en Chile. El laboratorio también publicó el primer informe anual del Repositorio<sup>81</sup>, que analizó detalladamente los 50 sistemas presentes en la plataforma.

En cuanto a sus actividades de divulgación, GobLab organizó 28 eventos propios que contaron con la participación de más de 3.200 personas. Además, tuvo presencia en otros 40 eventos organizados por diferentes instituciones, lo que equivale a un promedio de 1.3 eventos de divulgación por semana. Esta labor también se expandió a canales alternativos (o no estrictamente científicos y académicos) de comunicación, logrando 35 menciones explícitas en prensa, todas ellas con una valoración positiva. El valor publicitario equivalente de estas menciones se estimó en 60.704.182 CLP, lo que refleja el impacto y la favorable recepción de sus actividades en la sociedad y los medios de comunicación (Memorial GobLab 2022, p 6).

En el contexto del área estratégica 2 ( Integración de estándares éticos en el uso de datos en el sector público), el proyecto Algoritmos Éticos<sup>82</sup>, respaldado por una alianza público-privada con el apoyo de BID-Lab, logró llevar a cabo una serie de iniciativas en colaboración con organismos asociados: la División de Gobierno Digital del Ministerio Secretaría General de la Presidencia - Guía "Formulación Ética de Proyectos de Ciencias de Datos"<sup>83</sup>, destinada a los servidores públicos; con el Consejo para la Transparencia, se desarrolló y probó una propuesta de Instrucción General<sup>84</sup>, que se convertirá en la primera de su tipo en toda Latinoamérica, tras un proceso participativo; El Instituto de Previsión Social<sup>85</sup> - una solución basada en inteligencia artificial que beneficiaría a aproximadamente 300.000 personas que tienen derecho a recibir asistencia estatal pero aún no la han solicitado, entre muchos otros

Para finalizar este recorrido por las actividades de incidencia del laboratorio de innovación pública de la UAI, tenemos la tercera estrategia (Promover la creación de una Infraestructura de Datos Integrados (IDI) en Chile) donde en enero de 2022, se estableció un convenio de colaboración con una duración de 5 años para dar origen al "Consorcio IDI - Universidades por una Infraestructura de Datos Integrados". Este consorcio está compuesto por cuatro instituciones académicas líderes en Chile: la Universidad Adolfo Ibáñez, la Pontificia Universidad Católica de Chile, la Universidad de Chile y la Universidad Diego Portales. La

---

<sup>81</sup> [Primer informe anual del Repositorio](#)

<sup>82</sup> [Proyecto Algoritmos Éticos](#)

<sup>83</sup> [guía Formulación Ética de Proyectos de Ciencias de Datos](#)

<sup>84</sup> [proceso participativo para desarrollar y testear una propuesta](#)

<sup>85</sup> [se sumó como socio del proyecto](#)

creación de este consorcio resultó de una colaboración iniciada por GobLab UAI a finales de 2019 con el propósito de promover el uso seguro y accesible de datos administrativos integrados para investigadores y profesionales interesados en temas de relevancia pública, y en la formulación y evaluación de políticas públicas desde el ámbito gubernamental. La operación del Consorcio se lleva a cabo a través de la coordinación conjunta de diversas entidades, incluyendo la Escuela de Gobierno UAI, el Centro de Encuestas y Estudios Longitudinales de la UC, la Escuela de Gobierno UC, el Instituto Milenio Fundamentos de los Datos (IMFD) y el Centro de Estudios de Conflicto y Cohesión Social (COES).

Además, en marzo de 2022, la directora del GobLab concluyó su participación en el Comité Asesor Ministerial de Datos de Interés Público del Ministerio de Ciencia, el cual presentó un informe que delineó directrices para la gobernanza de datos en Chile<sup>86</sup>. Este informe incluye recomendaciones específicas para la creación de una infraestructura integral de datos en el país.

Percibimos entonces, a partir del análisis de las actividades de laboratorio en las cuales GobLab estuvo involucrado, que la incidencia es otro pilar fuerte de la institución. Tras siete años de existencia, el laboratorio ha crecido y potenciado sus capacidades y estrategia de actuación en formación, investigación e incidencia.

### ***6.3.7 Publicaciones relevantes ni siempre son artículos: el texto y sus múltiples funcionalidades***

Otra dimensión importante que GobLab, siendo parte de una estructura universitaria, también desarrolla es el de las publicaciones. A través de una rápida visita a su sitio web, al clicar “qué hacemos” también es posible navegar a “publicaciones”, abajo de “Ética de datos”, línea de investigación auto declarada por el laboratorio.

Las 33 publicaciones atribuidas al laboratorio están anexadas en la misma página. Es de fácil acceso y cada una de ellas contiene una breve descripción del contenido de la referida publicación, y un botón vinculado a un enlace para su descarga.

Hay gran variedad de temáticas, algunas escritas en inglés, pero la mayoría disponible en español. Entre artículos, informes, “estados de la evidencia”, guías y manuales, al mirar rápidamente derecho, algoritmos y transparencia, parecen ser los términos más utilizados. No logré identificar todas las publicaciones de los colaboradores del GobLab. La imagen a

---

<sup>86</sup> [Informe con lineamientos para una gobernanza de datos en Chile](#)

continuación resume las palabras presentes en los títulos de dichas publicaciones. Los términos en mayor tamaño representan las palabras de mayor frecuencia y peso. “Algorítmica”, “transparencia”, “pública”, “algoritmos”, “ética” y “estado”, son los términos que más se repiten en su base de publicaciones.

**Figura 46.** Palabras más frecuentes- Publicaciones GobLab.



### ***6.3.8 Internacionalización como prerrogativa de existencia: marcas de una práctica mundializada***

Así como el laboratorio desarrolla una forma de trabajo de forma bastante asociativa a través de su amplia red de colaboradores, GobLab posee una agenda bastante internacionalizada. No sólo profesores y colaboradores poseen nacionalidades y currículum contruidos en diversos países del mundo, sino que también busca expandir sus redes de influencia alrededor de latinoamérica particularmente.

En el ámbito de la internacionalización de la formación, durante 2022 se observa el fortalecimiento de su colaboración iniciada en 2021 con la Escuela de Gobierno y Transformación Pública del Instituto Tecnológico de Monterrey, México. Además de la realización conjunta del curso "Inspecciones proactivas con ciencia de datos", se logró poner

en marcha el Diplomado Internacional en Ciencia de Datos para la Dirección Pública, con una primera cohorte de 19 estudiantes que comenzaron en agosto de 2022.

Asimismo, GobLab continuó su colaboración con otros laboratorios de innovación pública en América Latina a través de la Red InnoLabs, a la cual se ha unido desde 2018. Por primera vez, actuó como anfitrión del Encuentro Anual de la Red, que fue organizado en colaboración con el Laboratorio de Gobierno de Chile. En este contexto, se llevó a cabo el primer seminario internacional de la Red, que contó con la participación de nueve laboratorios miembros de Chile, Argentina, Brasil, Colombia, Costa Rica, España, México y Portugal, con un total de 51 investigadores.

Adicionalmente, se estableció una colaboración internacional con la ONG ILDA y la Universidad del Rosario de Colombia, con el propósito de promover una alianza público-privada a nivel regional para fortalecer las capacidades gubernamentales en la transparencia de los algoritmos utilizados en la toma de decisiones. En este contexto, se llevaron a cabo dos eventos dirigidos al sector público en abril de 2022, uno en colaboración con la División de Gobierno Digital y otro con el apoyo del Consejo para la Transparencia. Estos eventos permitieron entablar contacto con diversos gobiernos de la región para explorar su interés en avanzar en materia de transparencia algorítmica.

También la directora del GobLab UAI participó en el Working Group P3119 AI Public Procurement Standard del IEEE<sup>87</sup>. El Instituto de Ingenieros Eléctricos y Electrónicos es una asociación global de ingenieros dedicada a la normalización y desarrollo en áreas técnicas. Este grupo está trabajando en la formulación de un estándar internacional para las adquisiciones públicas de inteligencia artificial. En este contexto, se compartió la experiencia en la temática a partir del proyecto Algoritmos Éticos y se lideró el subgrupo encargado de definir el problema.

Al final del año 2022, GobLab fue invitado a participar como ponente en 21 eventos internacionales en 10 países distintos, que incluyeron a México, Reino Unido, Ecuador, Estados Unidos, Guatemala, Brasil, República Dominicana, Uruguay, Colombia y Canadá.

Como se pudo observar en los apartados anteriores, la categorización y presentación del laboratorio de innovación pública de la UAI posee una actividad bastante activa,

---

<sup>87</sup> <https://standards.ieee.org/>

configurándose como unos de los más importantes laboratorios en ciencias de datos, datos de gobierno y temáticas de demandas sociales.

Como también defendí en el inicio de este capítulo, los estudios de laboratorio son esenciales en el campo de las Ciencias, Tecnologías y Sociedad, ya que proporcionan una ventana única para comprender cómo se moldean y se aplican la ciencia y la tecnología en la sociedad. La etnografía en tal laboratorio promete enriquecer nuestro conocimiento sobre estas dinámicas, ofreciendo una perspectiva invaluable que puede contribuir al desarrollo de políticas públicas que tengan en su ADN una toma de decisiones más informada y a partir del contexto de la ciencia y la tecnología.

Esta investigación parte del entendimiento de que los laboratorios son lugares donde las normas culturales y las dinámicas sociales pueden influir en la toma de decisiones científicas y tecnológicas. Estudiar estas dinámicas proporciona una visión de cómo se desarrolla la ciencia y la tecnología en la sociedad a través también de la comprensión de las dinámicas sociales y culturales. Se comprende que los resultados de la investigación y la innovación en laboratorios pueden tener un impacto directo en las políticas públicas. Analizar estas interacciones es esencial para entender cómo se formulan y aplican políticas relacionadas con la ciencia y la tecnología, y por consecuencia, en ciertas dimensiones de la sociedad.

## Cap 7: La ciencia de datos es el nuevo big data

Durante los primeros años de investigación estuve “persiguiendo” los big data. Mi primera inquietud fue acercarme al fenómeno de las tecnologías asociadas y su utilización en la práctica investigativa de las ciencias sociales de Latinoamérica y el Caribe. A causa de esta motivación, emprendí los estudios bibliométricos en Scielo y Scopus para conocer las publicaciones científicas y la difusión del conocimiento científico sobre los datos masivos. Fue a partir de este movimiento que surge la necesidad de entrevistar a los científicos autores de los textos pertenecientes a las dos bases de datos, para poder profundizar en la comprensión del proceso de fabricación del conocimiento social a partir/con y a través de los macrodatos. En medio de este proceso, como se menciona, comienzo a trabajar en un laboratorio involucrado en temas de innovación, tecnología y grandes datos de gobierno, lo que me permitió tener un contacto aún más cercano con el campo que me propuse estudiar en la presente investigación.

Hoy, puedo decir que cuatro años fue un buen tiempo para presenciar, en primera línea y en vivo, las transformaciones que han experimentado los estudios de este campo. Primero, desde 2019 a 2023 se vuelve nítido un desinterés y desuso del término “big data” por parte de los investigadores vinculados al tema de tecnología e investigación social. Podemos verlo en el análisis de los estudios bibliométricos realizados: la cantidad de artículos que posee el término en los títulos, resúmenes o palabras claves comienza a disminuir en 2020. En Scielo, la mayor tasa de crecimiento de “publicación en big data” es entre 2016 y 2019 (el número de textos creció de 1 a 6 por año) y nunca más se observó tal crecimiento. En 2019 se mantuvo el número de 6 publicaciones al año, y en 2020 este número sube un poco más, pasando a 9 hasta el fin de este período.

Si hoy (en 2023) realizamos la misma búsqueda, se puede constatar que hay una tendencia de decrecimiento de las publicaciones. Si bien 2021 es el año que más se publicó sobre big data y ciencias sociales (11 publicaciones), vemos en 2022 una disminución significativa, con apenas 5 textos, cantidad equivalente a 2019, el periodo de mayor *boom* de la literatura académica especializada. En 2023, ya casi a fines de años, apenas 2 textos fueron indexados en Scielo, lo que corrobora la tendencia de disminución de las publicaciones que tiene big data como centro de la discusión en las ciencias sociales latinoamericanas.

En Scopus la realidad no es muy diferente, en la figura 36, vemos que en 2020 es el año con el mayor número de publicaciones sobre el tema de big data en el mundo. Desde 2021 hasta

la actualidad, la tasa de publicación de los textos académicos en esta materia sigue disminuyendo, lo cual se repite en China y EUA, países que poseen la mayor representación en publicaciones sobre el tema.

La tercera evidencia de que el término “big data” viene paulatinamente cayendo en desuso, se puede ver en una de las entrevistas realizadas al personal del GobLab. Si bien esta entrevista tenía un objetivo distinto, pude “descubrir” este primer hallazgo al preguntar, a uno de los colaboradores del laboratorio, sobre su historia de fundación y la importancia de los diplomados que ofrecen como estrategia para promocionar sus iniciativas de formación, escuché el siguiente comentario:

Entonces, empezamos con el *framing* de *big data* [sobre la historia de los diplomados]. Teníamos un programa educativo que se llamaba Big Data para Políticas Públicas. Y por 2 o 3 promociones la tuvimos. Pero es un *framing* que hemos abandonado. Cambiamos el nombre porque es un *framing* que yo creo que está un poco pasado, como ya antiguo porque no refleja más lo que se hace. Y bueno, porque además nosotros nos situamos en el sector público, y no todo el uso de datos que el sector público hace es *big data*, puede ser datos que no se considerarían *big data*. Entonces como que el *framing* de *big data* yo ya prácticamente no lo uso, pero en el fondo nosotros estamos situados en la innovación pública, la transformación del Estado, en donde nuestros principales proyectos casi todos son de datos. Por eso, ahora los diplomados son en *ciencias de datos* y no *big data*. (MPH, Entrevista 01/2023 - Chile, mi énfasis. Mi traducción.)

Este comentario hizo sonar una alarma de alerta en la investigación. Les comparto un pedazo de una nota de mi diario de campo, escrita a fines de enero de 2022:

Acabo de salir de una entrevista con [omitido]. Cuando explicaba sobre el proceso de formación de servidores públicos en los cursos y diplomados ofrecidos por el laboratorio, comentó que las primeras promociones de las formaciones partieron con el término *big data* y que pasado un tiempo, lo habían cambiado a ciencia de datos. Esto me llamó demasiada atención. Esta misma información aparece de manera sutil en las entrevistas que he hecho con los científicos de las publicaciones. Todavía no he encontrado un autor que haya de verdad trabajado con *big data*, *big data*. Hay un interés por el procesamiento automático y computacional de los datos disponibles, inclusive a partir de otros tipos de tecnologías y métodos. Ojo en las “Ciencia de Datos” (Cuaderno de campo personal, enero de 2022).

Me quedé con este comentario dando vuelta en la cabeza, y resolví hacer una rápida búsqueda comparativa en Google Trends<sup>88</sup> inmediatamente después de la entrevista. Abrí la página de la herramienta y en el campo de búsqueda escribí “big data”. Filtre la selección en “búsquedas página web”, ajusté el periodo de tiempo entre 2004 (era el mínimo posible) y el presente (marzo 2023). En la segunda casilla que permite comparar términos escribí “data science”, lo que formó una nueva gráfica, a continuación.

**Figura 47.** Evolución comparativa del número de búsquedas de big data y data science



*Nota.* La línea azul representa el término big data, y la roja el de data science.

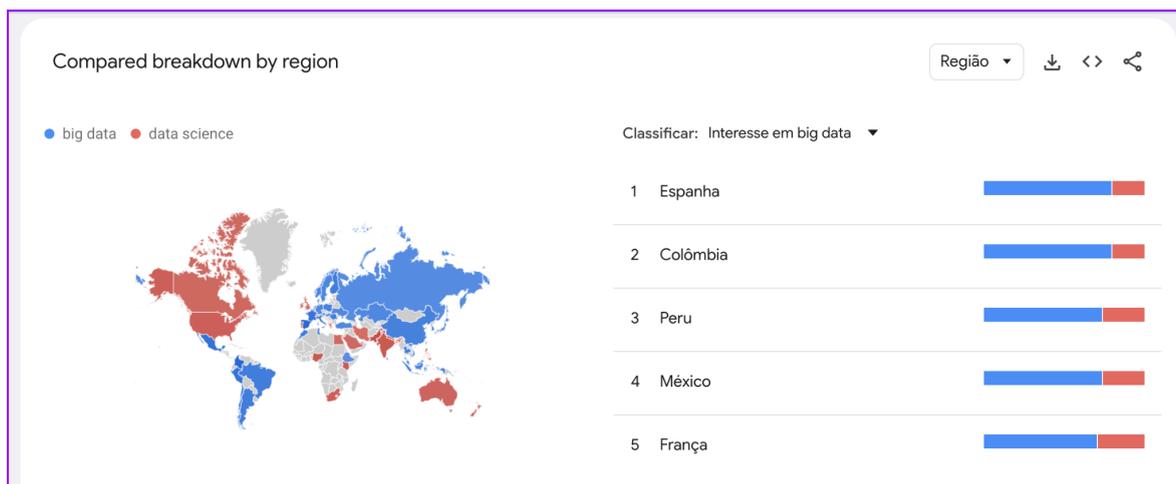
Como se puede observar, en las búsquedas en Google e internet, el interés de los usuarios cambia con relación a la frecuencia de los términos en cuestión en 2019. De 2004 hasta julio de 2011, ambos términos tienen prácticamente la misma ocurrencia de búsqueda, muy baja. A finales de 2011, el interés por el término big data crece de 5 a 14 búsquedas en internet. Esta tendencia sigue hasta enero de 2019, donde por primera vez se observa un cambio en el patrón de las búsquedas. El término data science pasa a ser más buscado en todo el mundo y

<sup>88</sup>Google Trends es una herramienta de Google que muestra los términos de búsqueda más populares en los últimos tiempos. La herramienta presenta gráficos que muestran la frecuencia con la que se busca un término concreto en varias regiones del mundo y en varios idiomas.. <https://trends.google.com/trends/>

el interés de las personas crece, al contrario del término big data, que desde entonces viene experimentando una baja de interés<sup>89</sup>.

Más adelante, en el mismo Google Trends podemos observar (Figura 48) el detallado de la información por región. Es interesante la distribución entre los diferentes términos de acuerdo con la división socioeconómica: El sur global, presenta una mayor concentración de búsquedas sobre temas relacionados a big data. Europa comparte el interés por el término pese a ser calificado como primer mundo. Sin embargo, el norte global presenta el mayor número de búsquedas por el término data science, junto a unos pocos países de África. Es más, los únicos 3 países africanos que presentan alguna búsqueda relacionada al tema que discutimos, se inclinan por data science.

**Figura 48.** *Búsqueda de los términos big data y data science pr región, desde 2004 a 2023*



*Nota:* Resultado búsqueda Google Trends.

Este panorama puede revelar similitud con lo que tradicionalmente sabemos de los estudios de ciencia, tecnología e innovación en el mundo. Tradicionalmente, los países del norte global son los que administran los medios, sistemas, recursos y diversos tipos de capitales para salir adelante en investigaciones, estructuración y consolidación de proyectos relacionados con tecnología (Orozco, 2021). Por lo general, es la región que presenta las tendencias tecnológicas e innovaciones y que involucran ciencia al resto del mundo.

<sup>89</sup> La réplica de esta búsqueda puede ser visualizada a través del siguiente enlace:  
<https://trends.google.com/trends/explore?date=all&q=big%20data,data%20science&hl=pt>

No sería extraño que una explicación para la preponderancia actual del término big data en Latinoamérica, Caribe y otros países del sur global, es su retraso a la hora de incorporar tecnologías, mientras que los países del norte:

Ya superaron la fase big data y entendieron la manera correcta de expresarse. Hacer justo a la cuestión. La ciencia de datos abarca cuestiones de big data y no también. En América latina es aún es *sexy* decir *big data*, ¡En gringolandia ya es vulgar! ¿me entendí? Los tipos estos ya tienen dominado la ciencia de datos y en unos años más qué nombres tendrán la evolución de esta disciplina, ¿verdad? Estamos siempre retrasados, es una pena, mas es la realidad. (RNN - Entrevista 41/2022 - Chile)

En Latinoamérica, todavía estamos descubriendo las promesas, impactos y desafíos teóricos que las tecnologías alrededor de los datos masivos traen a la investigación científica social.

Siguiendo con el análisis de la herramienta, Pakistán (31% big data - 69% data science), Nepal (32% - 68%), Bangladesh (35% - 65%), Filipina (35% - 65%), y India (36% - 64%), son los países que representan mayor interés por data science frente a big data, pasando a EUA, Canadá y Australia, países que tradicionalmente poseen vasta literatura en big data.

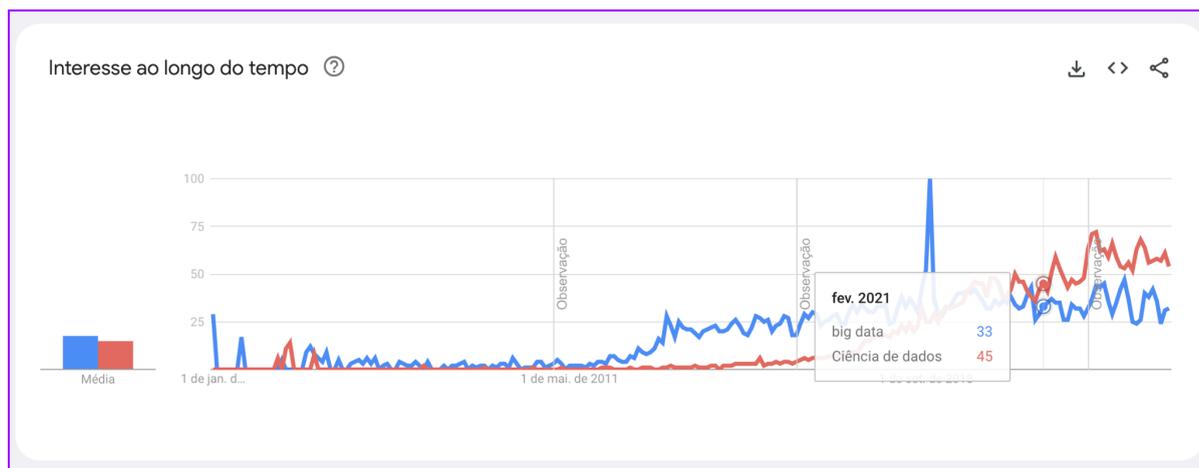
EUA ocupa la décima posición: 58% sobre data science contra 42% de búsquedas sobre big data; Australia, en la 11° posición de interés por data science disfruta de un porcentaje muy semejante: 56% - 44%; Canadá en 12°, con 45% (big data) 55% (data science). Con respecto a China, los datos no son considerados confiables por la ínfima cantidad de búsqueda y por el sistema de internet particular que tiene el país, por lo que no es posible capturar sus registros.

Con relación a los países de América Latina y el Caribe, regiones de nuestro interés, podemos ver en la tabla 23 que los países de esta región aún siguen interesados en el término big data. Chile, Argentina y Brasil son los países que menos búsquedas del término han hecho en la región en los últimos 2 años.

**Tabla 23.** Proporción de búsquedas relacionadas a big data y data science por país

País	% búsquedas Big data	% búsquedas Data Science
Colombia	80	20
Perú	74	26
México	74	26
Brasil	65	35
Argentina	63	37
Chile	62	38

Aunque en los países de Latinoamérica las búsquedas por “big data” sean más predominantes, la diferencia entre los términos viene paulatinamente disminuyendo, lo que podría representar una adecuación o incorporación de nuevos conceptos de forma tardía de su parte. En Brasil (figura 49), comienza la disminución en diciembre de 2021. En agosto de 2019, “data science” sobrepasa a “big data”. A partir de febrero de 2021 se ve un creciente interés por el término “data science”, continuando hasta hoy.

**Figura 49.** Relación entre los términos “big data” y “ciencia de datos” en Brasil, 2004 a 2023

En Argentina, “big data” viene ganando históricamente, y en enero de 2022, el término “data science” logra una ventaja significativa respecto a big data. En Chile, desde diciembre de 2020 data science sobrepasa a las búsquedas, manteniendo esta diferencia hasta hoy.

Resumiendo hasta acá, en América Latina, Brasil, Chile y Argentina, son los únicos países de la región que han incorporado este cambio reflejado en los parámetros de búsqueda de internet. Mientras que Colombia, Ecuador, México y Perú todavía exploran mayormente el término big data.

Con estos resultados en mano, pude confirmar la tendencia insinuada por mis entrevistados. Por ello, decidí prestar atención al tema durante las entrevistas, e incluso sumar preguntas sobre el trabajo de investigación en “ciencias de datos”, atendiendo a la nueva información.

A partir de entonces, pude constatar, sobre todo gracias a las entrevistas, que los pocos científicos y autores de artículos que efectivamente seguían trabajando el tema de sus publicaciones, también habían cambiado su vocabulario para referirse a éste. Ahora, se hacía presente la *ciencia de datos*, no sólo en cuanto disciplina, también se constituía como un paraguas conceptual para abarcar prácticas y tecnologías alrededor del procesamiento automático de datos computacionales, sean ellos masivos o no, como también la inteligencia artificial. Esto se tornó aún más evidente a partir de las entrevistas colectivas realizadas junto a Henry Chavez en su visita de trabajo a Chile<sup>90</sup>.

Entrevistamos a científicos autores de Scopus interesados en los macro datos en áreas distintas a las de las ciencias sociales. Al expandir el horizonte, nuestro contacto con los ingenieros y científicos de datos aumentó, lo que nos permitió establecer y entender más claramente las diferencias y complementariedades entre la ciencias de datos y nuestro tema de interés.

A partir de estas entrevistas, constatamos el nivel de involucramiento de estos científicos con el procesamiento de datos, independiente del tamaño de su base de datos. Ellos operaban aquella misteriosa caja de la limpieza, automatización y producción de la información de los proyectos de investigación donde participaban, pero no lo hacían solos. Descubrimos, entonces, la figura del “especialista”, término utilizado por la comunidad de ciencia de datos para referirse a los “no desarrolladores”, esto es, a los detentores de los *datos de entrada*<sup>91</sup>,

---

<sup>90</sup> A causa de su trabajo de investigación de post doctorado, vinculado al CTS-Lab Ecuador, Henry Chávez vino a Chile en 2022 para realizar algunas entrevistas a académicos interesados en temáticas big data. Durante su visita, entrevistamos a algunos investigadores en conjunto, con el propósito de reunir información para ambas investigaciones.

<sup>91</sup> Los datos de entradas son, de manera genérica, la base de datos disponible y suministrada por parte de los “contratantes” o interesados en la solución proveniente del procesamiento de datos. Pueden ser tablas de excel, un banco de imágenes satelitales u otro tipo, o bien una carpeta con miles de transcripciones de entrevistas o reuniones deliberativas, para dar algunos ejemplos. Están en su mayoría bajo la posesión de los especialistas, o responsables del proyecto a ser ejecutado.

contratantes o interesados en los análisis que serían generados a partir de de sus modelos y validación de los resultados generados.

Entonces pudimos ver, con más seguridad, que la gran mayoría de los colegas científicos sociales involucrados en proyectos de investigación sobre ciencia de datos no estaban, en efecto, trabajando con macrodatos pues no contaban con un volumen suficiente de datos para ser considerado como big data. Además, el hecho que los mismos autores operaban el procesamiento de la información generada, y no un “especialista”, marcaba una diferencia considerable.

Como se había detectado a partir del estudio bibliométrico y el análisis de contenido en la selección de la literatura estudiada, los textos traen, en su mayoría, un panorama muy superficial con respecto a la utilización, métodos e implicaciones de la implementación de tal tecnología a sus estudios. En general, o bien abordan la promesa que significa esta tecnología para las investigaciones en ciencias sociales (Silva, Bonacelli, y Pacheco, 2020; Qin, 2021), señalan el paradigma de trabajo (Borges, 2021; Moti Zwilling, 2023), refieren a las contribuciones más notables (Bragio, 2020; Lytras, Miltiadis y Visvizi, 2019), o a su impacto en la teoría social (Nascimento, 2016; Severiano, 2017; QIU et al 2017). Sin embargo, en Latinoamérica, prácticamente nadie en la academia, al menos en las bases de datos revisadas, Scopus y Scielo, está conduciendo un trabajo de investigación que efectivamente involucre la implementación y utilización de los datos masivos. El único grupo que de hecho estaría trabajando efectivamente con la utilización de datos masivos a partir de técnicas de procesamiento computacional de datos, era un grupo de investigadores que al momento de la entrevista no estaban más vinculados a la academia. Todos ellos trabajan de forma independiente, como consultores o “especialista en datos sociales” en proyectos en ciencia de datos que les iban apareciendo bajo demanda, sobre todo vinculados a la temática de operación de tránsito en grandes ciudades. Temática esta trabajada en una de las publicaciones que detectamos en la búsqueda bibliográfica (Sabino, A.-B., Reis-Martins, P., y Carranza-Infante, M. (2020).

Esta realidad se evidencia no sólo en el análisis del contenido de los textos académicos analizados, también se refleja en las conversaciones con los científicos autores de los artículos analizados para este estudio. Durante algunas de las entrevistas, se percibe una participación efectiva muy baja de parte de los autores con respecto al procesamiento de los datos (vamos hablar más de esto en el siguiente capítulo 8). Es decir, estos académicos

tendrían poca experiencia en trabajos investigativos que utilizan datos de tipo masivos. Cuando participan de investigaciones labeladas bajo big data, están, en realidad, en proyectos de ciencia de datos, y por lo general ejercen la típica función a la que hasta ahora se ha confinado al “trabajador de humanidades”, promoviendo o conduciendo técnicas de investigación social como lo son las entrevistas, encuestas exploratorias, estudios de satisfacción y opinión pública. Lo que hacen, en su mayoría, es coordinar y facilitar el proceso de entrega de un proyecto: Son el intermediario entre el objetivo del proyecto (lo que el contratante imaginó como solución efectiva) y el producto final que se entrega (ya sea un sistema, modelo algorítmico, plataforma o programa).

Durante una de las entrevistas, un ingeniero de *marching learning*, me explicó que un “modelo” era como el motor del auto, mientras que los científicos de datos vendrían a ser los responsables por garantizar el funcionamiento, abastecimiento y manutención de todo lo que conllevaba el funcionamiento de ese motor. Si esto es verdad, en mi opinión, los especialistas en humanidades vienen a cumplir el rol de un aceite, que suaviza y acomoda el funcionamiento de todo lo que hace el modelo, el proceso de fabricación de la información algorítmica.

Para un proyecto funcionar efectivamente, en el ambiente de la ciencia de datos, es necesario que sea funcional, lo que significa que debe estar conectado a las necesidades y problemas de los usuarios: sea él un ciudadano, un servidor público, o la policía del barrio. Un proyecto de naturaleza en ciencias de datos, va de la mano con la voluntad de incidir y “transformar, mejorar, optimizar, tornar eficiente, más rápido” alguna cuestión, aparato o proceso utilizado por un grupo social. Por lo tanto, su existencia depende de la aprobación de una comunidad de humanos. Es aquí donde entran los especialistas sociales: en el proceso de traducción entre el mundo real, las necesidades de los usuarios y las capacidades necesarias a desarrollar.

Como traductores, suelen ser especializados en otras áreas del conocimiento diferentes a las humanidades: ingenierías ambientales, ingenierías diversas, salud y medicina, derecho, economía, urbanismo y arquitectura social, geofísicas. No trabajan directamente con la manipulación de datos, más hacen puente entre estos dos mundos. Pueden ejercer una actividad semejante a los “analistas”, o el especialista, para usar una palabra nativa de los programadores/cientistas de datos, proporcionando respaldo y confiabilidad a los resultados encontrados a partir del trabajo del científico de datos.

Algunos de los científicos sociales mencionan haberse aproximado al tema de los datos masivos justamente para jugar el papel de consultor/especialista. Son incorporados a proyectos en ciencia de datos, y debido a naturaleza reflexiva de la profesión y su quehacer científico, durante el contacto que tienen con el mundo de las decisiones automatizadas y los sistemas computacionales vinculado a cuestiones sociales, se ven interpelados a reflexionar sobre las potencialidades, desafíos y engaños de la tecnología de investigación al cual fueron introducidos. El resultado de tal realidad es el preponderante número de artículos que tienen una característica exploratoria de terreno teórico: surgen a partir de la reflexión de los científicos sociales acerca de las posibilidades e implicaciones y mutuas afectaciones con relación a la propia disciplina, actuación, mercado de trabajos y desdoblamientos teóricos e incluso epistemológicos.

Analizando este panorama, y con algún tiempo transcurrido que me facilita la reflexión, surge, nuevamente, la obra de Knorr-Cetina (2005) sobre el proceso de modificación y artificialidad que implica la construcción de un texto académico al compararse con la realidad de los acontecimientos de los laboratorios:

“El vínculo entre el laboratorio y el artículo científico no puede establecerse por reglas de transformación cognitivas. [...]. El artículo científico es *producto* de ese doble modo de producción: no es su reflejo, resumen o descripción. El modo de producción instrumental que da como resultado las mediciones de laboratorio supone una descontextualización casi total, mitigada solamente por las fundamentaciones que se encuentran de las anotaciones de los científicas. El modo de producción literario cuyo producto es el artículo terminado ofrece una recontextualización, pero, como hemos visto, no una recontextualización que nos devuelva las memorias de trabajo de laboratorio. La transición es, al mismo tiempo, una conversión de las propias trazas escritas. Excepto en la memoria de aquellos que estaban presentes durante el proceso, se trata de una transición irreversible”. (Knorr-Cetina, 2005, p. 285)

Con el paso del tiempo mi propio trabajo de terreno se desenvolvía a la par de lo que percibía del estudio de los textos: yo misma, como investigadora y trabajadora social en proyectos de ciencias de datos, tampoco estaba operacionalizando estos datos. Más bien, estaba haciendo el viejo y conocido trabajo de las clásicas investigaciones sociales (cualitativas y cuantitativas). Excepto que junto con mis colegas de humanidades, éramos los más capaces de transitar y transmitir los mensajes e información de modo comprensible para ambos

frentes de trabajo, y así las actividades avanzaban dentro de las expectativas y las posibilidades de operación y contexto.

Gracias al trabajo en terreno en conjunto a la lectura de textos, que fui capaz de percibir que:

1. La ciencia de datos dió lugar al concepto de big data en contextos de trabajos e investigación vinculado al procesamiento computacional de datos masivos o no, y
2. Existen muy pocos científicos sociales involucrados en el proceso de manipulación de la información proveniente de big data y el procesamiento computacional de datos

Estas dos conclusiones conducen a las siguientes preguntas: Si los científicos no están trabajando o están involucrados directamente con la ciencia de datos, que es la nueva manera de definir conceptualmente trabajos que se utilizan de procesamiento computacional de datos sociales ¿qué están haciendo entonces? ¿por qué se justifica, o en base a qué se da el cambio entres estos términos? ¿sería algo ocasional, o habría una dimensión conceptual más evidente tras este cambio de términos?

Como vimos en los párrafos de arriba, los científicos sociales estudiados trabajan, en realidad, en un ambiente de investigación, o están insertados en un contexto laboral, donde el procesamiento computacional de datos es utilizado como un recurso para la obtención de colecta y análisis de datos sociales (mezclados a demás capas de información). Es importante destacar esta diferencia pues al fin y al cabo, antes de decir que el término big data ahora es *old fashioned*, ya que constatamos su desuso, fuimos experimentando, como comunidad científica vinculadas a tales temáticas, una maduración con respecto a la comprensión de las capacidades y límites de la tecnología big data, y por conseguinte su conceptualización.

Como se constató a partir de las entrevistas con los autores de las publicaciones en la fase de estudio bibliométrico, así como de mi propia experiencia en campo como investigadora y especialista en big data, la consolidación del término “ciencia de datos” en lugar de “big data” no refleja simplemente una tendencia pasajera, sino un avance teórico y metodológico significativo. Este cambio responde a una mayor comprensión de las técnicas computacionales de manejo de información, así como de sus implicaciones, recursos y consecuencias en la investigación. A lo largo de los años y con el desarrollo de esta investigación, se ha evidenciado una maduración en el campo científico y profesional que justifica un ajuste conceptual y terminológico para describir más adecuadamente proyectos de

esta índole. Este proceso de maduración en la comprensión y consolidación del campo se manifiesta claramente en la evolución de su nomenclatura.

En la siguiente sección, les presento una breve demostración de los principales motivos por los cuales la **ciencias de datos** se vuelve el nuevo objeto de análisis de esta investigación. Con el fin de adecuar mi investigación a la contingencia del proceso investigativo experimentado durante mi trabajo, se hace necesario que actualice los términos a sus debidos significados. Es decir, adaptarme a la realidad del contexto de la cual soy parte y con la cual trabajo. Busco, además, entender y evidenciar, a través de mi propio trabajo científico, por qué es importante el correcto uso del término y el concepto que involucra.

### **7.1 Big data versus ciencia de datos? diferenciando conceptos aparentemente similares**

Según algunos autores especializados, la diferencia fundamental entre "big data" y "ciencia de datos" radica en su enfoque y alcance teórico (Rahman, 2017; Sharma et al., 2021), mientras la ciencia de datos es una disciplina que se ocupa tanto de datos no estructurados como estructurados y abarca todo lo relacionado con la limpieza, planificación y análisis de datos, big data refiere al hecho de que un banco de datos sea de grandes cantidades, o variedad (Kitchin, 2014). También explicita que, según su opinión, los macrodatos aún no se pueden procesar eficientemente con las aplicaciones actuales, o esta capacidad está concentrada literalmente en manos de 2, 3 organizaciones en todo el mundo (Sharma et al., 2021).

En resumen, mientras que la ciencia de datos se ocupa de la limpieza, planificación y análisis de datos, el big data refiere a la cantidad de datos que se están generando, los cuales son demasiado masivos para ser procesados por las aplicaciones actuales accesibles. Por otro lado, a propósito de big data, los autores reconocen que el surgimiento de las nuevas tecnologías para capturar, comunicar, generar y conectar información son características distintivas de nuestra época, responsables por “la revolución Big Data”<sup>92</sup> y como consecuencia, han abierto el camino para el desarrollo de la ciencias de datos (Brady, 2019).

---

<sup>92</sup> Refiere a eventos ocurridos después del primer uso documentado del término “big data”. Aparece en un artículo de científicos de la NASA en 1997 describiendo el entonces problema de visualización de datos debido su masividad, lo que era un problema con la capacidad de la memoria principal, del disco local e incluso del disco remoto (Press, 2014).

Concluyen su raciocinio presentando algunas de las tecnologías asociadas a big data, como los sensores automatizados, lenguajes de programación, inteligencia artificial, telecomunicaciones, aprendizaje automático, redes sociales y el Internet, fueron artefactos que posibilitaron el surgimiento de una disciplina que utiliza varios métodos científicos (incluyendo estadísticas, minería de datos, aprendizaje automático, análisis de redes, visualización de datos, bases de datos y programación), para extraer información de grandes volúmenes de datos a través de modelos algorítmicos y procesos computacionales, algunas de las tecnologías que se utilizan en la Ciencia de Datos (Sharma et al., 2021; Brady, 2019; Rahman, 2017) ). Lo anterior puede verse resumido en la siguiente tabla.

**Tabla 24.** Comparación entre big data y ciencia de datos

Aspecto	Big Data	Ciencia de Datos
Enfoque principal	Gestión y procesamiento de grandes volúmenes de datos, a menudo de diversas fuentes y alta velocidad (5Vs).	Proceso completo de obtención de conocimiento a partir de datos, incluyendo recopilación, limpieza, análisis y modelado.
Técnicas y metodologías	Almacenamiento distribuido y procesamiento en paralelo, como Hadoop y sistemas NoSQL. Menos énfasis en técnicas estadísticas tradicionales.	Utiliza una variedad de técnicas, incluyendo estadísticas tradicionales, aprendizaje automático y minería de datos, seleccionando las apropiadas para el análisis según características de los datos disponibles.
Objetivos	Gestión de grandes conjuntos de datos y garantizar su disponibilidad y accesibilidad.	Aprovechar los datos disponibles para obtener información, tomar decisiones informadas, hacer predicciones, descubrir patrones y tomar decisiones informadas en diversas aplicaciones.
Contexto aplicación	Situaciones con gran velocidad y volumen de datos, como redes sociales, sensores IoT y análisis de registros de servidores; datos de Gobierno.	Amplia gama de contextos, se utiliza en una variedad de campos, desde la investigación científica hasta la toma de decisiones políticas, económicas y geográficas, hasta las empresariales como la detección de fraudes, personalización de recomendaciones y estudios de mercado.
Habilidades requeridas	Enfocadas en administración de sistemas de almacenamiento de datos a gran escala y gestión de infraestructuras como Hadoop.	Incluyen programación, estadísticas, visualización de datos, aprendizaje automático y habilidades de comunicación y análisis de datos.

En resumen, la principal diferencia entre big data y ciencia de datos radica en su enfoque. El primero se centra en la infraestructura y la gestión de grandes conjuntos de datos, mientras

que el segundo se enfoca en el proceso completo de análisis y obtención de conocimiento a partir de los datos, utilizando una variedad de técnicas y metodologías. La ciencia de datos es una disciplina más amplia que abarca la gestión de datos, pero también se preocupa por cómo esos datos se utilizan para obtener información significativa y tomar decisiones fundamentadas.

Observar cómo el lenguaje y los términos utilizados en el campo de las ciencias sociales y la investigación han evolucionado con el tiempo es comprender que la transición del término "big data" al de "data science" refleja una evolución en la forma en que se aborda y se percibe la investigación basada en datos. Más específicamente sobre la ciencia de datos, Brady, en su texto "The Challenge of Big Data and Data Science" (2019), presenta una tabla donde expone las principales 7 actividades relacionadas a la disciplina:

**Figura 50.** *Las siete actividades de las ciencias de datos según H.E Brady (2019)*

<b>Activities</b>	<b>Examples</b>
Data gathering, preparation, and exploration	Survey data, experimental data, genomic data, textual data, administrative data, image data, web data, and sensor data Data cleaning and exploratory data analysis methods for checking on outliers and data quality
Data representation and transformation	Relational and nonrelational databases Networks and graphs Other mathematical structures for data
Computing with data	R and Python Programming packages, text manipulation languages Cluster and cloud computing Reproducible workflows
Data modeling	Determining or hypothesizing data generating probability functions, structural and predictive modeling
Data visualization and presentation	Types of visualizations and graphs Rules for labeling and presenting data Psychological impacts of various displays
Data archiving, indexing, and search and data governance	Standards for open data and reproducibility Determining rules for access and privacy protection where necessary
Science about data science	How people do data science Impacts of data science and big data on society

*Nota.* Tabla extraída del artículo citado "The Challenge of Big Data and Data Science" (Brady, 2019).

Como se puede notar, la ciencia de datos abarca un mayor universo investigativo, incorpora métodos y otras maneras de operacionalizar diferentes tipos de datos y contextos de desarrollo. Mi defensa es que, con el pasar de los años y una mayor incorporación de tecnologías de ciencias de datos en el mundo científico social, nos uniremos a la tendencia mundial en desutilizar el término "big data" para referir a ambientes, contextos, sistemas o ecosistemas en ciencias de datos.

Esta orientación nos encamina a la posibilidad de hacer algunas observaciones que podrían ser relevantes para pensar los orígenes y consecuencias de esta transformación terminológica. El primero, sobre su potencial de alcance: la adopción del término ciencia de datos puede indicar una ampliación en el alcance de la disciplina y su permeabilidad o capacidad de conversar con diversas áreas del conocimiento. También transmite más confiabilidad y mejor adhesión por parte de los científicos, pues su diseño de trabajo involucra testeos de metodologías tradicionalmente más establecidas como la estadística, el aprendizaje automático y la visualización de datos, los que nos lleva a nuestro segundo punto: la naturaleza multi e interdisciplinaria que requieren las investigaciones en ciencia de datos.

Como se habló en la primera parte de este texto, la capacidad de digitalización de la información y su lectura a través de los códigos universales de los computadores permite transitar entre las disciplinas, métodos y teorías de modelamiento con mucha más practicidad, facilidad y velocidad, que nunca antes. Las bases de datos digitales unifican la producción de saberes que hasta muy poco tiempo atrás eran puestas como antagónicas y opuestas.

La incorporación de la ciencia de datos a las ciencias sociales, o vice y versa, subraya la creciente interdisciplinariedad de este campo no sólo por parte de las demandas de los proyectos, también por parte del movimiento de investigación. Los científicos sociales están adoptando enfoques y herramientas de análisis de datos más avanzados para abordar preguntas complejas y multidimensionales, tal como los ingenieros y tecnólogos están incorporando estudios de opinión, encuestas, etnografías y entrevistas para acercarse a las necesidades de sus clientes/usuarios, y también para cuestionarse sobre el sesgo y la falsa neutralidad que hasta entonces suponían algo inevitable y natural al oficio (Gualda, 2022).

A mi entender, esta transposición de términos sería fruto de los desafíos éticos y metodológicos asociados al procesamiento computacional de los estudios constituidos alrededor de grandes y variados datos. En consecuencia, a las ciencias sociales y las humanidades se nos abre un campo gigantesco por estudiar, explorar, definir y delimitar: los acuerdos y consensos, controversias y disensos sobre los desafíos en la investigación contemporánea y su impacto en la forma en que se realiza la investigación en nuestra disciplina.

Al parecer, finalmente mi investigación estaría en línea con una tendencia importante en las ciencias sociales, donde la transición de "big data" a "ciencia de datos" refleja una evolución en la forma en que se aborda la investigación basada en datos. Explorar el impacto de esta

transición en la investigación social contemporánea y los desafíos y recortes asociados a la sociedad contemporánea es un aporte valioso para los trabajos académicos de las ciencias sociales del presente y futuro.

## 7.2 El big data como una técnica presente en la ciencia de datos

He podido concluir que big data es la reunión de características asociadas a los datos digitales, que durante los años 2010 reaparecieron con una nueva fuente de datos que podría ser utilizada para entender la sociedad de una manera general. A mediados de dicha década, cobra interés en la academia, en la investigación teórica y práctica, expandiendo su territorio más allá del mundo del mercado, bancos y marketing. La industria también muestra que las competencias deseadas han pasado de habilidades de "manejo de big data" a habilidades en "ciencia de datos" y "análisis predictivo". Los informes anuales de empresas como IBM, LinkedIn, y Indeed reflejan este cambio, indicando una preferencia creciente por expertos en ciencia de datos sobre especialistas en big data. Con una breve revisión de estudios de tendencias laborales o reportes de LinkedIn ya se hace posible demostrar una perspectiva consolidada sobre este cambio en la demanda y denominación profesional, evidenciando cómo el término "ciencia de datos" ha absorbido al "big data" en la práctica y definición de competencias<sup>93</sup>.

Su ascensión, sin embargo, no es gratuita. Esto se debe a ciertas contingencias de la contemporaneidad. Big data nace digital. Es hijo del cambio al cual frecuentemente llamamos era. Vivimos en la era de la información: computadores, internet, digitalidad, datos digitales, lo que está íntimamente vinculado a la capacidad del big data de existir (White y Breckenridge, 2014). La expansión de la capacidad de procesamiento de los computadores, y del abaratamiento de su costo también es un gran responsable por este *boom*.

El aumento de la *capacidad computacional de vida*<sup>94</sup> además de permitir el almacenamiento de una infinidad de datos con relativa facilidad y conveniencia, facilitó el desarrollo de

---

<sup>93</sup> Esta información puede ser demostrada y para saber más:

<https://www.forbes.com/sites/louiscolumbus/2017/05/13/ibm-predicts-demand-for-data-scientists-will-soar-28-by-2020/>

<sup>94</sup> Es decir, la capacidad de producción de huellas digitales y datos de todo tipo: personales, de salud, de relación, de registro de actividades y compromisos, de geolocalización, de educación, de deudas, de ingresos. Es la capacidad, por un lado, de almacenamiento de datos y por otro, de desarrollar las tecnologías, materiales y soportes, que se requieren para sostener la materialidad de la vida, de dejar registros (con datos masivos), de finalmente, existir.

técnicas, herramientas, modelos matemáticos, y dispositivos para procesar y analizar. La capacidad de procesar un gran volumen de datos, de muchas variedades y con una velocidad bastante importante, hizo que esta tecnología se tomara un espacio en los laboratorios y programas de investigación científicos de diversas áreas del conocimiento, a tal punto de transformar su alrededor, primero, y a la definición de su propio concepto.

Sabemos que la implementación de una nueva técnica requiere de un periodo de apropiación, adaptación y tiempo de prueba para la absorción de sus potencialidades, límites e incluso una manera de incorporar la *narrativa del resultado*<sup>95</sup> que esta nueva técnica y metodología imprime en los frutos de una investigación (Kuhn, 1962).

Mi intuición es que conforme el big data se incorpora a la investigación, la comunidad científica se va dando cuenta de sus reales implementaciones e incluso limitaciones. Independiente de las justificaciones y razones para esto, es cierto que la creciente curiosidad por los datos masivos, ya habiendo pasado cinco años desde su descubrimiento en la ciencia, ha ganando nuevo nombre y apellido: ciencia de datos.

Con el tiempo, el acercamiento a las tecnologías big data<sup>96</sup> ha permitido entender mejor sobre lo que significa trabajar con datos masivos y de gran escala, y lo que conlleva un ambiente de procesamiento computacional. Parte de ello se debe a la maduración de la comunidad académica de las investigaciones en, para y con big data.

Acercándonos a este escenario, vemos que la definición de ciencia de datos se establece como el término más apropiado (Slota, S. C., et al, 2020). Ya no es difícil encontrar una buena definición sobre “ciencia de datos” en Google, ni saber lo que es un algoritmo o la existencia de la inteligencia artificial, por ejemplo. De hecho, es casi imposible ser un humano conectado a internet sin tener una idea del vocabulario y los conceptos tecnológicos asociados, los cuales eran inimaginables hasta apenas tres años atrás: inteligencia artificial, inteligencia generativa, algoritmos, procesos automatizados, entre otros.

Podríamos decir, entonces, que sociedad y laboratorio maduraron en conjunto la incorporación de estos conceptos a la vida cotidiana y laboral. Al empezar el doctorado tardé

---

<sup>95</sup> Es decir, el proceso que da origen, como un molde, a los resultados alcanzados. Este molde tiene una manera de proceder, esconde o no evidencia las negociaciones y decisiones entre la capacidad humana, la máquina, los recursos y habilidades disponibles a su momento. Contiene historia y forma en los resultados, que aparecen en mejor o mayor nivel de destreza, según sean estructurados estos factores.

<sup>96</sup>El entorno tecnológico que involucra datos masivos incluye: infraestructura de datos, capacidad computacional para procesar, capital humano que sepa manipular los datos digitales, el idioma y el saber que es necesario para operar en conjunto con los tiempos y problemas involucrados.

dos años en encontrar un congreso con una mesa exclusiva sobre big data. Ahora, hace dos semanas, recibo un correo de invitación para participar de la Conferencia Internacional BigSurv (Big Data Meets Survey Science), donde el título llamativo y *cool*, usa el término “big data”, pero el cuerpo del correo utiliza “ciencia de datos” cinco veces, como es posible constatar en la figura abajo.

**Figura 51.** *Email invitación participación BigSurvey*

Estimada/o colega:  
 Como usted conoce, la **Conferencia Internacional BigSurv (Big Data Meets Survey Science)** se realizará por primera vez en Latinoamérica, en la ciudad de Quito, del 26 al 29 de octubre de 2023. Se trata de una iniciativa global y multidisciplinaria de altísimo nivel, que junta a la academia, la industria y la sociedad civil, a nivel nacional e internacional.

Como podrá ver en el siguiente enlace, destacados científicos y profesionales del mundo debatirán en BigSurv sobre los temas más innovadores en la ciencia de los datos y las encuestas, así como las ciencias computacionales aplicadas a las ciencias sociales: <https://www.bigsurv.org/program23>

Además, en el marco de BigSurv se realizarán **dos eventos vibrantes y complementarios**:  
 El primer evento es el **Data Challenge** sobre el tema de la desnutrición crónica infantil, conjuntamente con la Fundación REDNI ([www.redni.org](http://www.redni.org)). Este tema nos debe convocar a todos en el país, y queremos que también los jóvenes participen en este empeño. El objetivo del Data Challenge es usar la ciencia de los datos para el bien común, y están convocados profesionales, estudiantes, investigadores interesados en los objetivos del desafío. Además, los ganadores recibirán importantes premios. Toda la información sobre este Data Challenge se encuentra en el siguiente enlace: <https://www.bigsurv.org/DataChallenge>

El segundo evento son los  **cursos cortos de actualización**. Toda la información, puede encontrarla aquí: <https://www.bigsurv.org/shortcourses>

El registro on-line termina el día **15 de octubre**, por lo cual le solicitamos amablemente que comparta esta información con los profesionales y jóvenes que usted considere que estén interesados en BigSurv, los cursos cortos y el data Challenge.

Atentamente,  
 Ana Lucía Córdova Cazar, PhD  
 Presidenta del Comité Científico de BigSurv23

---



Ana Lucía Córdova Cazar, PhD  
 Profesora  
 School of Business

Universidad San Francisco de Quito  
 T: (+593) 2 297-1700 ext. 2125  
 Correo: [alcordova@usfq.edu.ec](mailto:alcordova@usfq.edu.ec)  
 Diego de Robles y Vía Interoceánica, Quito, Ecuador  
<http://www.usfq.edu.ec>

Nota de descargo: La información contenida en éste e-mail es confidencial y sólo puede ser utilizada por el individuo o la institución a la cual está dirigido. Esta información no debe ser distribuida ni copiada total o parcialmente por ningún medio sin la autorización de la USFQ. La institución no asume responsabilidad sobre información, opiniones o criterios contenidos en este mail que no estén relacionados con asuntos oficiales de nuestra institución. Disclaimer: The information in this e-mail is confidential and intended only for the use of the person or institution to which it is addressed. This information is considered provisional and referential; it cannot be totally or partially distributed nor copied by any media without authorization from USFQ. The institution does not assume responsibility about the information, opinions or criteria in this e-mail.

*Nota:* Captura de pantalla del correo de invitación

Se puede concluir que lo que está en juego, finalmente, no es el big data en sí, sino el sistema, el ambiente que hay a su alrededor. Pues es a partir de la necesidad de construir y comprender este ambiente digital/computacional asociado a los grandes datos digitales, que el término big data se iría metamorfoseando hacia una ciencia que extrae información y soluciones de un volumen considerable de datos digitales: la ciencia de datos. Además, el término big data se ha vuelto un poco ambiguo, ya que no hay una definición clara que limite qué es "masivo" en términos de cantidad de datos.

Por lo anterior, hablar de ciencia de datos es más apropiado llevando en consideración que tal conceptualización involucra una serie de dimensiones propias de la constitución de un campo disciplinar científico tales como la capacidad de operacionalizar y incluir diferentes tipos de información digital que requiere de una serie de aparatos, saberes, lenguaje, parámetros y procedimientos, que van mucho más allá de que un dato sea grande (Brady, 2019). Ordenar la base de datos, procesar, testear, analizar (o sea, hacer cosas) con los datos disponibles requiere de una logística semejante a la que tiene un laboratorio tradicional de biología molecular, por ejemplo. No se estudian los efectos de una nueva vacuna para un tipo de

gripe si no hay microscopios y tubos de ensayo. De igual modo, no existe una análisis de la sociedad a partir de grandes datos digitales sin un *data center*, sin una GPU<sup>97</sup>, un profesional de TI, horas de programación y una capacidad de procesamiento y operación.

La ciencia de datos requiere una estructura material importante: una arquitectura de cables, supercomputadores, aire acondicionado, miles de Kilowatt-hora de energía y capital humano cualificado para funcionar. Este último, es el papel de otro tipo de profesional muy asociado a la ciencia de datos, el ingeniero de *machine learning*<sup>98</sup>. La persona que garantiza el funcionamiento que conlleva “rodar/correr” (utilizando lenguaje nativo) esta infinidad de datos. Aparte de la problemática de infraestructura, está el saber sobre la gobernanza de datos, qué involucra el proceso de decidir a quienes pertenecen, cómo protegerlos, qué hacer con la información, y cuáles aspectos incide la utilización de estos datos.

Están los matemáticos e ingenieros trabajando en modelos y ecuaciones de lógica que van por detrás de los algoritmos, en caso de utilizar grandes datos y automatización de procesos. Están los programadores, construyendo la fundación de cada sistema, aportando tijolos hechos de datos, parámetros y pequeñas maniobras que permiten que letras y números se transformen en botones clicables y aplicaciones amigables a un usuario común: los desarrolladores de *front end*. Están también los responsables por el proceso de visualización e interpretación de los datos, lo que varía según cada metodología de procesamiento adoptada, experticia en el tema y el conocimiento de programas, sistemas y lenguaje computacional. Los que ayudan en el proceso de cobrar vida a los datos, transformar las “salidas” en informaciones relevantes y aptas para la utilización de los especialistas y usuarios.

Como se puede ver, se requiere de un equipo variado e interconectado, aunque cada uno trabaja a partir de un saber bien específico. Sus procedimientos son afines a los laboratorios tradicionales, hay ciertos métodos, orden de hacer las cosas, secuencias de programación y cronología del desarrollo. Todo siempre orientado a una pregunta: ¿estamos resolviendo lo que nos propusimos resolver?

Así, parte del hallazgo de esta investigación es que el término big data potencializó o fue responsable por introducir la manera actual de investigar ciencia de datos. El ambiente investigativo computacional, lo que culmina en la ciencia de datos, es más amplio que big data, pues involucra más complejidad para existir y operar, y sin duda está en plena

---

<sup>97</sup> Ver glosario

<sup>98</sup> Para conocer la definición de esta profesión, ver glosario.

ascensión, en especial a partir de 2021, hasta la actualidad (2023). En resumen, la creciente popularidad del concepto de ciencia de datos y consecuente decadencia del uso del término big data refleja una comprensión más amplia y sofisticada de cómo se usan los datos para tomar decisiones informadas y para mejorar sus capacidades de rendir, sea en un ambiente investigativo, o en empresariales y de gestión de datos. O sea, el término "big data" surge principalmente para describir un fenómeno de volumen y complejidad masivos en datos. Sin embargo, a medida que las técnicas y herramientas de análisis de datos han avanzado, la "ciencia de datos" emerge como una disciplina en sí misma, que involucra no solo la manipulación de grandes volúmenes de datos, sino una metodología integral que abarca estadística, aprendizaje automático, y otras ciencias computacionales.

En el capítulo que sigue, presentaré con más detalles la vida de este laboratorio en ciencia de datos e innovación de gobierno, así como el trabajo etnográfico. A partir de la descripción de sus actividades laborales, será posible percibir, de mejor manera, la real conformación conceptual que les acabo de presentar, bien como comprender las consecuencias de esta profesionalización alrededor de las informaciones computacionales vinculadas al estudios sociales. De igual manera, tiene como objetivo demostrar y entender las especificidades, prácticas, tecnologías usadas, flujos, controversias, consensos y particularidades derivadas del trabajo investigativo construido a su interior.

## Capítulo 8: Investigaciones en un laboratorio en ciencias de datos

Los resultados que presentaré, vienen de mi experiencia como investigadora asociada al laboratorio de innovación de Gobierno de la UAI (GobLab), durante 1 año y 4 meses de permanencia como profesional contratada. Durante este periodo, tuve la oportunidad de acompañar la ejecución de tres proyectos en ciencias de datos vinculados a dicho laboratorio. También es resultado de mi participación como miembro del comité del Repositorio de Algoritmos Públicos de Chile, y de las más de 40 entrevistas realizadas con expertos en investigación en big data y ciencia de datos en Brasil y Chile, principalmente entre 2020 y parte de 2023.

Los hallazgos a presentar son fruto de la observación en mi recorrido etnográfico. Si bien trabajar de cerca con proyectos en ciencias de datos, como investigadora social, no fue una estrategia metodológica inicialmente planeada, fue gracias a la decisión de abordar este proyecto etnográficamente lo que posibilitó armar esta suerte de *mapa-esquema* de las prácticas socio-técnicas (Latour, 1992; 2008; 2013). Estas prácticas las pude vivenciar y analizar desde ambos puntos de vista: mi participación en el laboratorio y mi rol como científica interesada en los estudios de la ciencia, tecnología y sociedad.

De esta inserción en terreno y observación de los actores presentes, tanto humanos como no humanos, logré identificar por lo menos seis implicaciones metodológicas diferentes en cuanto a la forma en que se constituyen y generan los conocimientos en un laboratorio con tales características.

Antes de llegar a estos hallazgos, les presentaré un esquema que trata de ilustrar el flujo, contexto y la conformación del laboratorio estudiado, con la finalidad de presentarles un panorama general del funcionamiento de sus actividades de investigación y producción de conocimiento, incidencia, innovación y soluciones tecnológicas a la comunidad la cual está dirigida.

Hemos discutido que el desarrollo de las tecnologías digitales de las dos últimas décadas ha dado lugar a la generación de una enorme cantidad de datos y, como consecuencia, nuevas herramientas para tratarlos y servirse de ellos. El acelerado desarrollo de las tecnologías de la información y comunicación (TICs), así como el proceso de digitalización de todas las actividades humanas y no humanas en las dos últimas décadas han incrementado exponencialmente las capacidades de la sociedad moderna para producir, compartir y utilizar todo tipo de datos. Tal realidad acaba por transformar diversos aspectos de la vida contemporánea, incluida, por supuesto, la ciencia (Castells, 2010; Schwab, 2016).

Generados por millones de dispositivos electrónicos e interacciones en línea, estos datos requieren ser procesados y analizados. Producto de esta necesidad, a mediados de los años 2000, surgió, sin precedentes, una serie de conceptos, técnicas y herramientas asociadas a lo que varios expertos y medios de comunicación denominaron como *big data*. Este marco conceptual y tecnológico permite gestionar la producción masiva de datos y utilización para la generación de conocimiento, valor e innovación.

Su rápida difusión y el despliegue infraestructural que lo ha acompañado, como vimos en el capítulo anterior, da lugar al nacimiento de otro término asociado a este universo: la ciencia de datos, que a su vez, permite la configuración de una suerte de ensamble sociotécnico (Latour, 2007; 2008) global de datos e información al que se integran miles de organizaciones, investigadores y profesionales de las más distintas áreas.

La investigación basada en ciencia de datos gana espacio a partir de la promesa de permitir a los científicos comprender mejor las tendencias y patrones de sus diversos objetos de interés, mejorar la toma de decisiones u obtener ventajas competitivas a través del procesamiento computacional de los datos, tras la intención de optimizar el proceso de análisis de datos variados, desestructurados y de gran volumen, en cortos periodos de tiempo, y sin dejar de plantear importantes desafíos en términos de privacidad, seguridad, regulación y ética (Brady, 2019).

El uso de la "ciencia de datos", ese campo de estudio que se enfoca en los principios, métodos y técnicas para extraer información y conocimiento a partir de datos complejos (Provost et al., 2013) junto con herramientas y técnicas de big data, viene ganando un espacio creciente en la academia y en las iniciativas de innovación tecnológica. Este avance es especialmente notable en las diferentes esferas gubernamentales de los países, los cuales, a su vez, están cada vez más comprometidos con generar proyectos de ciencias sociales centrados

en el análisis computacional de extensos conjuntos de datos gubernamentales, especialmente en América Latina.

De la interacción entre la investigación académica y las iniciativas en innovación tecnológica por el bien común, se evidencia, como una especie de efecto colateral, la emergencia y necesidad de mejorar y adaptar los procesos metodológicos tras el uso de datos masivos para la resolución de problemas. Esta optimización colateral, ocurre gracias al grado de experimentación metodológica que conlleva el procesamiento computacional de grandes datos, por un lado, y por otro, debido a la naturaleza curiosa y experimental atribuida a los científicos y académicos partícipes de estos tipos de proyectos (White y Breckenridge, 2014).

En mi inserción etnográfica pude, además, entender la relación del laboratorio con los resultados obtenidos; es decir, cómo es el proceso en que se generan sus saberes. Paralelamente, a partir de un segundo nivel de producción de información, observé cómo la artesanía de la construcción de estos saberes, incide en el estructurar, el proceder, y el mantener un tipo peculiar de producción de conocimiento.

Con esto en mente, consideré importante estudiar el desarrollo de las capacidades operativas y metodológicas de la investigación de los sujetos involucrados (científicos y no científicos) y el contexto espacial de producción. Con ello, me interesa mostrar el espacio de un laboratorio de investigación en ciencia de datos y cómo definen su estructura y funcionamiento. También elucidar sobre los sujetos involucrados en este espacio (humanos y no humanos), sus funciones, roles profesionales y agencia investigativa. Finalmente, les presentaré las implicaciones metodológicas para la investigación social a partir de las ciencias de datos observadas durante mi estudio.

El trabajo etnográfico permitió interactuar directamente con el mundo de las investigaciones en ciencias de datos, tanto con investigadores de ciencias “duras” como de las ciencias sociales, cuyas prácticas giran en torno a esta intersección entre la dimensión técnica e infraestructural de la tecnología y los problemas, controversias y cuestionamientos sociales, éticos y políticos que derivan de su implementación.

La siguiente figura intenta sintetizar y esquematizar la dinámica, las dimensiones y actores fundamentales que participan de esta interacción en base a las observaciones etnográficas realizadas en el GobLab, entre 2021 y 2022.



trabajada. Como bien lo sintetiza uno de los expertos en ciencias de datos contratado por GobLab, ingeniero en computación, post doctorado en formación:

Yo trabajo en muchos proyectos [respondiendo sobre las especificidades de su trabajo]. Yo soy experto en procesamiento de lenguaje natural, imágenes, *computer vision*, etc. Siempre que me llaman a un proyecto yo le digo “listo,ok. Dígame la *data* de entrada, dígame la *data* de salida y qué es lo que se quiere”. ¿Me entiende? Es imposible rodar un proyecto sin saber la *data* que disponen. El tono del proyecto proviene de esta *data* de entrada, y sabiendo lo que se pretende hacer, uno va eligiendo los modelos que trabajar. Y si la *data* de entrada no está así tiki-taca, hay muchos algoritmos que te ayudan a interpretar la entrada, muchos. (RNN - Entrevista 41/2022 - Chile).

Testimonios como este se repiten varias veces tanto durante el trabajo etnográfico, como en las entrevistas realizadas, lo que evidencia la íntima relación que existe entre las esferas de datos y la de problemas, Es necesario explicitar que alrededor de la esfera de los datos (E1) orbitan otros dos aspectos muy importantes para la ejecución de los proyectos de investigación: **la capacidad de infraestructura** (Cp3) y **la capacidad de gestión y gobernanza de datos** (Cp4).

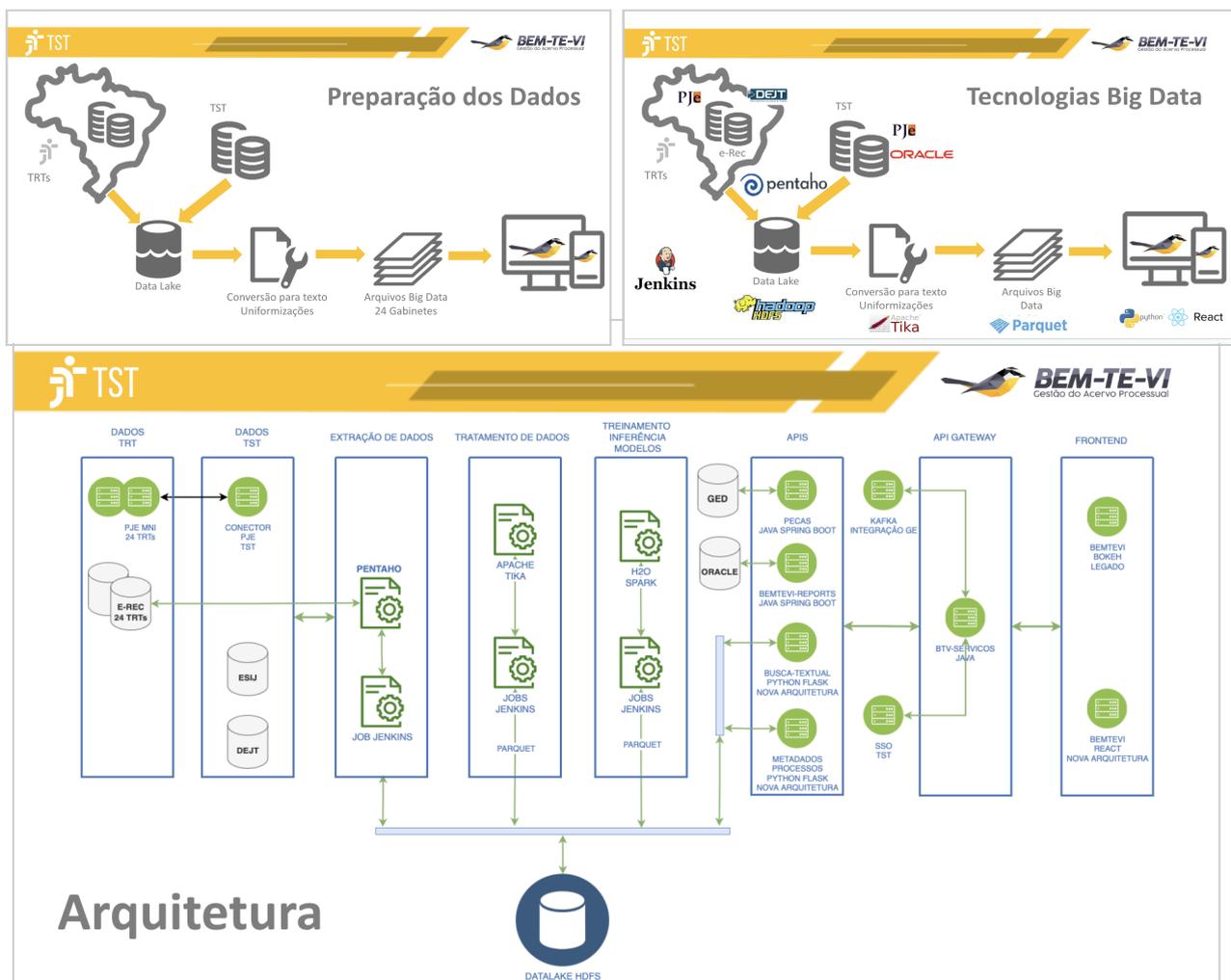
Esto, porque la capacidad de infraestructura computacional (Cp3) se determina por la calidad, las características y la potencia del material computacional con el que contamos (procesadores, memoria, almacenamiento, nube, etc.), es decir, la arquitectura con la que se ensamblan dichos componentes y los sistemas informáticos (softwares) de los que disponemos para poder operarlos. Una capacidad infraestructural adecuada es esencial para procesar grandes cantidades de datos y llevar a cabo las complejas tareas de análisis y modelado. Entre estas tareas se incluyen la capacidad de realizar consultas y análisis de datos, la capacidad de almacenamiento de toda información que ya se posee y la que va ser generada y la capacidad de procesamiento en tiempo real, entre varios otros aspectos. Uno de los más importante es la capacidad humana, que se refiere a la presencia de profesionales capaces de operar y hacer funcionar la logística de datos, desde su obtención, hasta el análisis y la visualización. También es importante contar con personal capacitado en la gestión de proyectos y en la toma de decisiones, para que el proyecto pueda avanzar más allá del procesamiento de los datos.

Además, la Cp3 incluye la compra de equipos y computadoras de alto rendimiento, softwares y herramientas de análisis de datos, así como la configuración de sistemas de almacenamiento y redes de comunicación para garantizar la disponibilidad y seguridad de los datos. Todo ello es necesario para una efectiva operación de un sistema de investigación en ciencia de datos.

La efectividad de un sistema refiere a la capacidad de un sistema para realizar operaciones de cálculo y procesamiento de datos. Esto incluye aspectos como la velocidad del procesador, la cantidad de memoria RAM, el almacenamiento en disco.

En el siguiente conjunto de figuras, vemos las diferentes etapas y el ambiente propicio para la implementación de una infraestructura computacional en un proyecto en ciencia de datos a partir del uso tecnologías big data e inteligencia artificial, el cual tuve la oportunidad de visitar en Brasil. Esta figura es parte del material de difusión del sistema *Bem-te-vi*: Gestión de acervo procesual del Tribunal Superior del Trabajo de la República de Brasil.

**Figura 53. Infraestructura de Operación Sistema Bem-te-vi, 2023**



Nota. Extraído de [Sistema Bem-te-vi](#) Accedido en octubre de 2023.

Acá vemos la decena de aplicaciones, etapas y tecnologías necesarias para el procesamiento, implementación y manutención de un sistema aparentemente simple. En este caso, ejecuta la categorización de más de 5 millones entre piezas jurídicas y acuerdos de los 24 gabinetes de los ministerios del Tribunal Superior del Trabajo de Brasil. El procesamiento de toda esta información y la estructura que requiere, fue parte de una iniciativa interna de la institución, con el fin de implementar efectivamente un “modelo de lenguaje jurídico trabalhista” que pueda organizar, clasificar y sugerir agrupamientos de temáticas semejantes a los millones de procesos laborales en el Tribunal Superior del Trabajo de Brasil. De ese modo, las actividades judiciales de la institución ganan eficacia.

Por otro lado, la capacidad de gestión y gobernanza de datos (Cp4), determinada por las competencias exclusivamente humanas, organizacionales y de regulación que se encargan de dar buen uso a la capacidad infraestructural para recolectar, almacenar, procesar, analizar, preservar, eliminar y asegurar la integridad y disponibilidad de los datos antes, durante y después de un proyecto de investigación de este tipo. La figura 54 abajo sintetiza los diferente procesos y aspectos partícipes de Cp4:

**Figura 54.** Línea del tiempo de la capacidad de gestión y gobernanza de datos



Nota. Fuente: revista *Une de la normalización Española*, 2023.

La Columna A de la figura arriba, evidencia una situación de simbiosis entre estas dos diferentes capacidades. La capacidad de infraestructura computacional, como podemos percibir, está contenida entre unas de las primeras responsabilidades atribuidas a las capacidades de Gobernanza de datos. La columna B, es una invitación a “humanizar”

(utilizando una expresión muy evocada en las conversaciones entre los diferentes especialistas) el proceso de investigación a partir del procesamiento computacional de los datos. Es la parte considera “de humanas” donde, normalmente, los gestores académicos y contractantes responsables del proyecto, orientan las actividades en términos de aporte y relevancia al grupo o comunidad a ser beneficiada, o también es el momento dedicado a pensar sobre los riesgos, limitaciones y desafíos a serem superados por parte de todo el equipo de investigación.

La columna C es la comuna más específica con respecto a gobernanza. Es el momento de definir los líderes de cada equipo. También está en juego la observación de los procesos de toma de decisión como cuando aparecen controversias respecto a las metodologías utilizadas, o la necesidad de elegir una camino frente el otro.

Me gustaría saber como ustedes toman las decisiones dentro del equipo técnico. Imagina que están ahí entrenando un modelo y existen 2 posibles tipos de modelos que podrían funcionar perfectamente con los datos que ustedes están trabajando. Imagina que el tiempo es limitado, tienen 2 semanas para hacer todo y terminar el proyecto. ¿Con quienes conversaron sobre y en su opinión, quienes toman las decisiones finales?

Entrevistado: Ah. Ok. Pero no es una respuesta fácil, ya le digo. Esto va depender de cada proyecto y básicamente de la expertise y el cargo del responsable técnico. Y claro, del día, de la hora, de la disponibilidad del momento y de mi voluntad! [risas] Pero hablando en serio. que tienes que entender, que muchas de las cosas que se hace en un determiname como momento del internamiento del modelo y de hacer correr la info es que es feeling. Es experiencia y confianza en mi propia capacidad de pensar y raciocinar junto con la máquina. De verdad que no sabría decir que tenemos un protocolo de acción para esto. Normalmente yo estoy ahí en el computador, conversando con la máquina y los datos, si todo fluye bien y hace sentido en mi cabeza, yo sigo no más. Pueden pasar semanas sin que yo no necesite de la opinión de nadie, hasta que...Un problema me sobrepasa. No se que hacer. Ahí pido opinión a alguien, la persona de más fácil acceso que yo creo que va saber lo que yo necesito saber...Y ahí conversamos, hago test y veo si es esto o no. Hasta ahora sólo me ha tocado 1 vez que tuvimos que hacer una reunión colectiva para decidir como arreglar la falla de un aspecto del modelo. Pero el problema no era el modelo...Llegamos a un acuerdo y listo. Tampoco hay tanto rencor en estar correcto o no, si uno fue el responsable por decidir aquello. al final en peor de los casos culpamos a la maquina! (JPG - Entrevista 09/2021 - Chile)

Como se ve, el proceso de gobernanza de datos en los ambientes de investigación en ciencia de datos es muy fluido. Se sabe, a través también de las entrevistas, que la preocupación con el proceso de acompañamiento de las decisiones, gobernanza y gestión de la información producida al interior de estos laboratorios aun es muy reciente. Esta discusión deriva del contacto con la emergencia de los temas de seguridad, privacidad y gestión de datos, impulsada por la implementación de mecanismos legales de transparencia y privacidad de datos, en los distintos gobiernos y bajo el proceso de modernización de cada Estado. La discusión sobre la LGPD (Ley General de Protección de Datos) de todo el mundo, ha empujado la preocupación con la gestión y gobernanza de datos al interior de iniciativas y actividades en ciencia de datos y datos digitales.

La columna D habla sobre el proceso de retroalimentación y evaluación de la efectividad y docencia de esos proyectos, su contacto con los usuarios finales y la resonancia de los resultados y soluciones generadas a partir del contexto de la investigación. Juntas, las dos capacidades, orbitan alrededor de los datos. La estructuración de un buen proyecto de investigación con estas características, necesita de una sólida constitución de dichas capacidades. Abajo (tabla 25), una síntesis de las competencias y aspectos que dan cuenta de la distinción entre estos dos tipos de capacidad, en un laboratorio de investigación en ciencia de datos con vocación hacia el público y lo social.

**Tabla 25.** *Capacidad infra computacional versus capacidad de gestión y gobernanza de datos*

Aspecto	Capacidad de Infraestructura Computacional (Cp3)	Capacidad de Gestión y Gobernanza de Datos (Cp4)	Aspecto
<b>Hardware</b>	Descripción del hardware utilizado, incluyendo servidores, estaciones de trabajo, dispositivos de recopilación de datos, sensores, etc.	Descripción de cómo se garantiza que los datos recopilados sean precisos, completos y confiables.	<b>Calidad de Datos</b>
<b>Software</b>	Resumen de las aplicaciones y plataformas utilizadas para análisis de datos, visualización, comunicación, gestión de proyectos, etc.	Aseguramiento de que los datos se mantienen íntegros durante su recopilación, almacenamiento y procesamiento.	<b>Integridad de Datos</b>
<b>Almacenamiento de Datos</b>	Cómo se almacenan y gestionan los datos, ya sea en servidores locales o en la nube, y cómo se asegura la integridad y seguridad de los datos.	Políticas y procedimientos para garantizar el uso ético y legal de los datos, así como la definición de roles y responsabilidades relacionados con los datos	<b>Gobernanza de Datos</b>

<b>Redes y Conectividad</b>	Descripción de cómo se conectan y comunican los dispositivos y sistemas en el laboratorio, y cómo se garantiza la disponibilidad y la velocidad de la red.	Medidas de seguridad y prácticas de privacidad para proteger la información sensible y cumplir con regulaciones de protección de datos.	<b>Privacidad y Seguridad de Datos</b>
<b>Seguridad Informática</b>	Medidas de seguridad, como cortafuegos, sistemas de autenticación y encriptación, implementadas para proteger la infraestructura y los datos.	Gestión del acceso a los datos y la compartición de datos con partes externas si es relevante.	<b>Acceso y Compartición de Datos</b>
<b>Recursos Humanos</b>	Disponibilidad de personal capacitado para gestionar y mantener la infraestructura.	Cómo se utilizan los datos para informar la toma de decisiones y cómo se fomenta una cultura de toma de decisiones basada en evidencia	<b>Toma de Decisiones Basadas en Datos</b>
		Si se proporciona formación a los miembros del laboratorio sobre la importancia de la gestión y gobernanza de datos.	<b>Formación y Concienciación</b>

La articulación de estas dos capacidades determina en gran medida los resultados y las soluciones ofrecidas por los proyectos de investigación, además del funcionamiento de la práctica investigativa<sup>99</sup>. En teoría, a mayor capacidad, mejor y más robustos serán los resultados y soluciones presentadas al final del proceso investigativo. Es importante resaltar que en este sistema existe una relación íntima entre capacidades humanas y no humanas. La máquina no “corre” los datos sin los comandos y algoritmos diseñados por los humanos, pero

<sup>99</sup> Nota campo - Brasilia - Feb. 2022 - Día 3: “Se uno supiera de la estructura necesaria para hacer funcionar estos micros click diarios que damos cotidianamente a través de nuestros computadores quedaríamos espantados. Hoy en una de mis visitas a laboratorios de ciencia de datos asociados a diferentes tribunales federales, en Brasilia, tuve por fin el coraje de pedir para conocer el tan famoso *data center* del Senado. No pude sacar fotos, ni grabar nada. Fue a la antigua: cuaderno y lápiz bolígrafo en la mano! Un montón de palabras inteligibles brotaban de mis anotaciones. Al final decidí sólo estar atenta. Desde el sistema de seguridad del data center, hasta su sistema de alerta y combate a incendio, los robots y computadores necesarios para armonizar “zillones de bytes” de información de toda *data* federal en Brasil, y dos veces, por si no sufren un ataque o pérdida! Ver los cables azules y verdes oliva pasando arriba mio como corrientes marítimas listas a desaguar en las gigantes cajas negras, cada de centros de computadores raros, traídos desde China y Estados Unidos. Millones de millares de dólares invertidos en equipos y profesionales a servicio “solamente” de la manutención, integridad y buen funcionamiento de la “casa de los datos de Brasil”, fue para mi toda una experiencia. Conocer el mayor data center de Brasil, instalado en el subsuelo de unos de los edificios de la explanada central en Brasilia. Lugar de carne y hueso, igual al ingeniero de sistemas que amablemente se candidató como mi guía. Hoy más que nunca me doy cuenta de la materialidad por detrás de la operación de los grandes datos, de las mega informaciones. Conversando con el Zé ahora en la noche [amigo del pregrado y magister, quien me abrigó en su casa] terminando concluimos que Marx, el viejo Karl barbudo de siempre, más una vez tenía razón: la alienación de todo trabajo, esfuerzo, capacidad cerebral y de recurso responsable por movimentar esta gigantesca estructura en todo mundo, el invisibilizada por la “magia por detrás” del funcionamiento de una tecnología. Sea una máquina a vapor en el siglo XVIII, o ahora en 2022, era posmoderna, seres de la revolución digital/virtual, a partir de la emergencia de la internet y derivados!”

estos tampoco pueden procesar datos masivos y complejos a una gran velocidad y asertividad. Uno sin el otro, no pueden existir; en conjunto, son responsables por promover la usabilidad de la data de entrada del proyecto, y deben ser orquestados entre las dos partes anteriormente mencionadas. Para que los problemas (E2) derivados de la esfera de datos (E1) puedan ser efectivamente contemplados, y los objetivos de la investigación alcanzados, se requiere la conformación de una triada de posiciones claves, que en conjunto, conforman una red de actores no muy fáciles de hacer confluír.

Acoplados a la esfera del problema (E2), dictando la conformación del diseño de investigación, están el *gestor del proyecto* (GP), el *experto técnico* (ExTec) y el *académico gestor* (AcG). A través de sus trabajos en conjunto confieren e ingenian la viabilidad del proyecto. En conjunto, esfera 1, y capacidad 3 y 4, son las responsables de la determinación de la esfera del problema (E2), estructurando así, las demás decisiones que tangecian esta esfera.

## **8.2 Esfera Problema (E2): la resolución de problemas como objetivo principal de los proyectos**

Una vez con los datos de entrada en manos, y con el conocimiento de la mínima infraestructura y gestión de gobernanza de los datos requerida, el equipo responsable por el proyecto empieza a trabajar sobre la construcción del problema a ser abordado.

En el caso del GobLab, siempre estuvo conformada por una serie de profesionales, al menos tres, variando conforme sus expertises y disponibilidades de trabajo. En efecto, el patrón de por lo menos 3 cargos/funciones<sup>100</sup> necesarios (GP + AcG + ExTec) se corrobora en la entrevistas. En conjunto componen el “cuadro de profesionales” responsables por la formulación del problema y la determinación de trabajo investigativo a ser construido. Son también los encargados de ir seleccionando a los demás profesionales necesarios para la ejecución del proyecto, respetando siempre afinidades en términos de habilidades técnicas y expertise en el tema, como es el caso de la contratación de los diversos tipos de especialistas

---

<sup>100</sup> Importante señalar que los nombres oficiales de los cargos de la investigación, una vez aprobado el convenio de colaboración o la firma que adjudique la ejecución de los proyectos, varía conforme las características de cada proyecto. Estos cargos son definidos también en función del modelo de exigencia de la parte auspiciadora o contratante y sus demás vínculos institucionales. Esta caracterización, por lo tanto, habla respecto al equipo mínimo posible de profesionales para lograr “levantar la ejecución de un proyecto” (para usar un expresión nativa) de investigación con las características que estamos presentando.

requeridos (en ciencias sociales, u otras áreas como consultores, científicos de datos, desarrolladores, asistentes, etc.)

El cargo de GP, por lo general, lo cumplen funcionarios públicos, servidores o comisionados provenientes del servicio contratante del proyecto (en su mayoría jefes de un departamento o división de innovación o TI de la organización). Es la figura clave, responsable por la gestión y el *pipeline*<sup>101</sup> de los proyectos. Esta persona experimenta la "vivencia del problema" al estar inmersa en su entorno de trabajo, y, gracias a su experiencia profesional y su contacto directo con los desafíos tangibles relacionados con la generación de datos, impulsará transformaciones y querrá incidir en su entorno, a partir de la base de información que dispone.

En el contexto específico del GobLab, el gestor suele ser un ex alumno de alguno de los programas de formación en ciencias de datos que tienen. Su profundo conocimiento de la base de datos y su familiaridad con los problemas cotidianos de su área de servicio lo capacitan para identificar los posibles problemas a tratar. Además, es quien comienza a diseñar el sistema que guiará el desarrollo del proyecto. El GP también asume la responsabilidad de identificar a los profesionales más adecuados para construir y ejecutar la investigación de manera efectiva. Su gran aporte es que refleja la importancia de tener a alguien en estrecho contacto con la realidad del problema, y que esté altamente motivado para utilizar la información en beneficio de la innovación.

Por otro lado, encontramos la figura del académico gestor (AcG), un profesional altamente calificado y experimentado en la dirección de proyectos relacionados con ciencias de datos. Este individuo desempeña un papel crucial al actuar como un puente entre el mundo académico y el ámbito de las ciencias sociales, es decir, es quien guía el diálogo entre la investigación y la implementación práctica.

Además, de ser el encargado de buscar fuentes de financiamiento a través de medios no convencionales, el profesional académico es solicitado en esta fase (E2) por su capacidad para conectar a los expertos y académicos pertinentes, de acuerdo a las necesidades del proyecto. Asume la responsabilidad de facilitar la comunicación y colaboración entre los especialistas en tecnología de la información y ciencias de datos que poseen conocimientos técnicos sólidos. Asimismo, acaba siendo el responsable por hacer el puente entre los saberes

---

<sup>101</sup> Carta Gantt del proyecto. Responsable por la planificación y diseño de cada etapa del proceso de la investigación.

técnicos más duros (profesionales de TI y ciencias de datos) con las demás necesidades del proyecto. Es por ello que su papel es crucial en asegurar una sinergia efectiva entre el mundo académico y el ámbito práctico del proyecto.

Por último, pero no menos importante, se encuentra la figura del experto técnico (ExTec), que no necesariamente será el científico de datos o desarrollador. Más bien, suele ser un ingeniero que comprende las particularidades y características de la implementación de proyectos en ciencia de datos. En muchas ocasiones, actúa como consultor externo, proporcionando orientación actualizada sobre la viabilidad del proyecto y los posibles enfoques metodológicos en ciencia de datos que podrían aplicarse. El experto técnico también ofrece directrices sobre las posibles vías de procesamiento de datos y puede contribuir al proceso inicial de garantizar la calidad de los datos de entrada. Es decir, aporta ideas y mecanismos para mejorar o adaptar los datos de entrada antes de que sean utilizados por los profesionales involucrados.

En este contexto de negociación, el ExTec generalmente trabaja junto con el ingeniero de machine learning (si es que esta figura existe en el proyecto) para determinar o asesorar sobre la estructura mínima necesaria para que el proyecto pueda llevarse a cabo. Su enfoque está centrado en los aspectos más técnicos y ejerce una influencia directa en el ámbito de los datos, como se refleja en la figura triangular en el esquema de la figura 52, ubicada justo debajo del círculo que representa los datos (CD).

Una vez que se ha conformado este trío de actores y se han establecido los problemas y objetivos del proyecto de investigación que se desarrollará, trabajan conjuntamente para definir la línea de acción del proyecto. Se dividen las tareas y, una vez que se ha cerrado el acuerdo de colaboración y se ha oficializado la investigación aplicada mediante la firma de un contrato o convenio, se incorporan al equipo los demás miembros.

Entre ellos se encuentran los *especialistas (Esp 5)*, expertos en la temática específica a ser abordada en el proyecto. Estos pueden ser profesionales en ciencias sociales o especialistas en otros campos relacionados con la problemática de la investigación. Su papel es crucial principalmente en la fase inicial del proyecto, donde se ocupan de la entrada de datos y la calidad de estos. También son relevantes en la fase final de la investigación (ETeo10), cuando se procede a la validación de los resultados y la implementación de las soluciones.

Los especialistas mantienen un contacto directo tanto con la esfera de problema (E2) como con el GP y el equipo de gestión de proyectos. La fluidez de la información que se intercambia entre los diferentes especialistas y los científicos de datos se convierte en una relación fundamental para el éxito del proyecto. La relación entre la esfera de los datos, la esfera del problema, los científicos de datos y los diferentes especialistas (E1 + E2 + CD + 5) es la base de toda la producción de información generada por el estudio. Tanto los especialistas como los científicos de datos son responsables de retroalimentar directamente el sistema, cada uno contribuyendo con sus conocimientos y procesos investigativos de manera complementaria.

Los especialistas, sin importar cuál sea su enfoque disciplinario, tienen la responsabilidad de enriquecer la discusión en torno al problema. Aunque no necesariamente sean los poseedores de la base de datos inicial o de los datos de entrada, su función es anclar la información en la realidad del mundo real y adaptar los datos para su uso en términos prácticos y de los parámetros necesarios. Son los encargados de determinar qué variables o tipos de datos son relevantes para el procesamiento de la información y la recopilación de datos. También asumen la responsabilidad de revisar los resultados y validar las soluciones frente a los diversos equipos del proyecto. Utilizan sus conocimientos previos como un mecanismo de conexión con la realidad, actuando como un recurso para seleccionar y determinar la pertinencia de las condiciones que se aplican a la investigación. Más adelante en este texto, profundizaremos en detalle sobre las aplicaciones y las implicaciones de cada uno de los tipos de profesionales involucrados en este sistema.

Por último, conectado a la E2, se encuentra el *consultor (C)*, quien, a diferencia de los especialistas, brinda servicios específicos de asesoramiento y evaluación en relación con el trabajo a realizar. Su rol es proporcionar orientación y recomendaciones cuando se requiere una especialización o una consulta específica sobre un tema que no es completamente dominado por los especialistas ya involucrados en el proyecto.

## **8.2 Espacio-Tema (ETem 6): unificando los diversos intereses, saberes y soluciones**

La interacción entre la esfera de datos (E1) y la esfera de los problemas (E2) da lugar a lo que el esquema (figura 52) identifica como el *espacio-tema* (ETem6). Se centra en la temática que aborda el problema, como por ejemplo, el mejoramiento del programa de asistencia

pública penal en el proyecto con la DPP, la prevención del fraude en el proceso de emisión de servicios de salud pública en Chile, o la preservación ambiental a partir de imágenes satelitales en el caso del proyecto Fondef SAMSARA.

El interés por investigar a profundidad un tema no es la prioridad de trabajos de investigación aplicada en ciencia de datos. El tema de investigación es más bien una especie de paraguas que abriga y hace convivir los datos, los problemas a ser resueltos y los diferentes especialistas involucrados. Los datos y la problemática a ser investigada, están delimitados por un espacio conceptualmente enmarcado a partir del tema adyacente al contenido de ambas esferas (Espacio-Tema 6 = E1 + E2).

Por lo tanto, el problema u objetivo de investigación, en la gran mayoría de los casos, está orientado a la incidencia y resolución de problemas a partir del conjunto de datos al que se tiene acceso. Todo lo demás en la investigación se adapta a la condición de los datos. El tema, demarca un territorio de intersección entre diversas prácticas investigativas, elecciones de metodologías, formas de trabajo, alineamiento de expectativas y priorización de actividades. Es en este espacio (ETem6) que se da el intercambio y co-dependencia laboral y de saberes. Es el lugar de convivencia íntima entre los datos y los especialistas, conformando así el motor de la investigación *métodos* () en la figura.

El ETem6 es también el territorio de los métodos y de las soluciones metodológicas encontradas, en relación a la dupla datos-problema. Consiste en la elección, por parte del científico de datos, de los mejores modelos y algoritmos que auxilian el proceso de procesamiento computacional de los datos y posteriores soluciones. Es el lugar de la construcción de los modelos, el testeo de las variables. Es donde los computadores y sistemas computacionales corren, a la vez que el proceso de producción del conocimiento empieza a tomar forma. Los resultados provienen de este espacio de producción.

### **8.3 Espacio-teoría (ETeo10): los enmarcado teóricos como mecanismo de validación de los resultados**

Al otro extremo del esquema, tenemos el *espacio-teoría* (ETeo10), donde se relacionan la producción de los resultados (7), su validación (8) y la posterior solución (9). Los *resultados* provienen del procesamiento computacional (automático o no) de los datos de entrada. Se

originan a través del entrenamiento de la máquina en conjunción con los modelos elegidos por el científico de datos, la información de la base de datos y las selecciones previas acordadas entre los distintos especialistas. A pesar de que estos resultados se originan en la interacción entre los científicos de datos (CD) y la esfera de datos (E1), en sí mismos no constituyen información relevante hasta que son sometidos al criterio y las pruebas de los especialistas. Este proceso de *validación de la información* (8) es fundamental para determinar precisión y posterior utilización de los resultados.

En esta etapa (8), se hacen evidentes las discrepancias y debates que surgen dentro del ámbito de la investigación a implementar. Aquí, la capacidad de comunicación y la habilidad para traducir los resultados a diversas esferas del proyecto son cruciales. Es fundamental que los expertos técnicos, así como los especialistas en ciencias sociales y otras disciplinas (5), estén preparados para intercambiar conocimientos, discutir posibles problemas relacionados con los resultados y, en última instancia, lograr un entendimiento mutuo.

En esta fase (ETeo10) se pone a prueba la habilidad del científico de datos (CD) para explicar por qué se elige un modelo específico o determinada metodología de aproximación a los datos, en lugar de otro. También se evalúa la capacidad de los especialistas (Esp 5) en interpretar los resultados generados por la máquina y traspasarlas a los líderes del proyecto, para que así, a través de un consenso sobre la validez de dichos resultados, la investigación pueda seguir adelante. La efectividad de los modelos y las soluciones propuestas depende en gran medida de estos acuerdos, y son fundamentales para las primeras pruebas de implementación de las soluciones propuestas (9).

Por lo general, he observado que la palabra final sobre la secuencia y permisión para la ejecución de la solución viene de los especialistas (5). Ellos son los responsables de trazar semejanzas y discordancias entre los resultados y la realidad observada y vinculado al entorno estudiado. Es en este momento que las teorías, o el saber conjetural de los especialistas son accionados, es decir, se incorporan al proyecto, como un recurso estratégico que visa validar o rechazar determinados resultados encontrados, como lo demuestra el siguiente fragmento de entrevista a un científico de datos vinculado al laboratorio:

\*Las frases *rosadas*, evidencian el papel del especialista, mientras que las *verdes*, del científico de datos.

“Bueno usted me dice “yo tengo un *data set* donde se calcula la discriminación”. Y a partir de estos [escribe en la pizara, refiriendose a la columna de las variables], del sexo, de la edad, de la región que habita, de cosas, de personas... puede ser cualquiera, no me importa el problema [referido al tipo de investigación].

Entonces, ¿dónde está la articulación cientista de datos y la Bárbara [el especialista]? Simple: yo sé lo que hay acá adentro, cómo trabajarlo, cómo parametrizarlo [con respecto a la esfera datos E1 - el procesamiento], pero Bárbara [la especialista] entiende muy bien las entradas y las salidas de estos problemas.

Pues yo le digo, “Bárbara, ¿tú quieres predecir la discriminación a partir de estas variables?”. Entonces lo primero es ¿qué variables de entrada tengo? Tengo 5000 [dice el especialista]. Entonces yo te digo “¿trabajamos con las 5000 de entrada?”, entonces tú dices “no, de estas 5000, sólo sirven estas 3000”, porque tú eres la conocedora, tú eres la primera que me da esa aproximación al modelo, para que al modelo no le cueste tanto entrenarse. Entonces tú dices “no, es que estas variables no se tienen, esta no da mucha información”. Y de ahí vamos calibrando el modelo.

Entonces tú como experta juegas en la entrada del modelo. Luego, en la salida. Entonces, en salida ¿cómo juega el experto? “Oye, me dijo 1.7 [el resultado del procesamiento] , ¿para ti está bien?”. Tú, Bárbara, como experta dices “sí, el 1.7 está bien para ese fenómeno. Lo veo con frecuencia así”. “ah bueno. Entonces usaremos el 1.7 para predecir la discriminación en este contexto de ahora en adelante”. Y por ahí va la cosa hasta que me des mas datos y tengamos que recalibrar el modelo y así infinitamente...” (RNN - Entrevista 41/2022 - Chile)

Si bien constatamos la existencia y funcionamiento de una colaboración recíproca entre humanos y máquinas, esta no es completamente simétrica, con respecto a los humanos, lo que también se puede corroborar en la cita de arriba. Hay una diferencia significativa en la interacción de los diferentes actores humanos y sus funciones en estos proyectos de investigación. Esto se evidencia en el hecho que los científicos de datos e ingenieros (CD) están en mayor contacto con los datos (E1) que con el problema (E2).

En cambio, los especialistas (Esp 5) se relacionan más con el segundo (E2), facilitando la elección de variables, validando correspondencias y proporcionando criterios de validez de los resultados. Ellos están detrás de la concepción y diseño de la investigación de manera global. Sin embargo, estos especialistas (Esp 5) no necesariamente “ponen las manos en la masa” del procesamiento computacional de los datos. Por lo general, ellos no dominan las técnicas y metodologías para manipular los datos, como lo hacen los expertos en datos y machine learning (CD). Estos últimos ejecutan y fiscalizan el proceso de “pensamiento de la

máquina”, coordinan comandos, buscan alternativas para procesar mayores cantidades de datos y hacen sugerencias de aspectos relevantes que deben incorporarse en el proceso de construcción de las bases de datos, lo que ellos llaman *data quality* (la calidad de los datos).

Posteriormente, los resultados obtenidos por la máquina deben ser validados por los especialistas en un nuevo territorio en el que confluyen ellos junto con las teorías y paradigmas de análisis conocidos para entender, interpretar o dar sentido a dichos resultados (ETeo10). La teoría ayuda a comprender el significado y la relevancia de los resultados, y a identificar sus implicaciones. También permite contextualizar los resultados dentro de un marco más amplio de investigación. No obstante, es importante resaltar que nuestras observaciones sugieren que no existe una fundamentación teórica del problema, sino un análisis teórico de los resultados.

Esta constatación, de que la teoría tiene una utilización más bien estratégica y táctica, que elucidatoria, es muy similar al proceso de alienación entre la teoría y la práctica definida por Knorr-Cetina (2005) cuando explica al lector algunos de los paradigmas de los estudios constructivistas de la ciencia:

En lugar de la conocida alienación entre la teoría y la práctica encontramos una maraña de acción/cognición a la cual ya no se le puede aplicar adecuadamente la idea preconcebida de lo que es una teoría. Según los propios científicos, en la investigación las teorías son más afines a políticas que a credos." Esas políticas combinan la interpretación con el cálculo estratégico y táctico, y son sostenidas por proyecciones metodológicas de! "cómo-hacer". Igual que la preocupación por hacer que las cosas funcionen, las políticas están necesariamente atadas a una estructura de interés. La teoría pura, entonces, puede ser calificada de una ilusión que las ciencias han conservado de la filosofía (p. 60)

La teoría puede ayudar a los científicos de datos a comprender el significado y la relevancia de los resultados, y a identificar las implicaciones teóricas de los mismos. Además, sirve para contextualizar los resultados dentro de un marco más amplio de investigación. Al tener un conocimiento sobre la teoría y los fundamentos teóricos, los científicos de datos pueden garantizar que sus resultados sean significativos y relevantes, teniendo implicaciones importantes para la investigación y para la toma de decisiones.

Con los resultados en mano, la intervención de los especialistas que circundan el espacio-tema (ETem 6) y la esfera problema (E2) es vital. Son ellos los que validan, en gran medida, los resultados observados por la máquina. Si son validados, los científicos de datos

vuelven a trabajar para conseguir una solución apta y amigable que contemple las necesidades acordadas en el proyecto. En esta fase, la colaboración entre los distintos especialistas y los científicos de datos es esencial. En este momento los diversos saberes teóricos y prácticos provenientes de distintas disciplinas convergen y culminan en la producción de la información final, necesaria para la validación de los resultados generados por la máquina y sus operadores. Después de esto, el equipo llega a un consenso, y, si están satisfechos, presentan la solución o producto prometido al principio. Es igualmente necesario resaltar que no siempre las soluciones a las que se llegan son las mismas que se propuso al inicio del proyecto. Pero, si el resultado final es satisfactorio y permite mejoras de algún proceso, casi siempre son devueltas a su entorno en forma de solución. Finalmente, es hora de que el científico de datos (CD) y el equipo técnico del proyecto (desarrolladores, programadores de *back* y *front-end*<sup>102</sup>, y demás técnicos en TI) empiecen a desarrollar las soluciones en base a los resultados y modelos provenientes del procesamiento de datos.

Después de este proceso, que pueden tardar de semanas a meses, se construyen entonces, las ***soluciones o productos*** (9), que pueden ser desde un sistema que auxilia a los funcionarios públicos en su servicio de atención a los ciudadanos, como un sistema interno de una institución capaz de detectar con anticipación riesgos de deforestamiento u otros inconvenientes ambientales. También, para citar una experiencia que conocí personalmente en una de mis vistas de campo en Brasil: un software capaz de optimizar el proceso de categorización de accordons judiciares, que inicialmente objetivaba facilitar el proceso de confección de decretos por parte de ministros de distintas incumbencias, en Brasil (uno de los brazos del proyecto Ulysses de la cámara de los Diputados Federales de Brasil)<sup>103</sup>.

La solución, o producto desarrollado (9), es generalmente es implementando desde el GP, con el objetivo de incidir en el contexto generado. Se lanza una versión de prueba o testeo, iniciando así la ***fase de implementación y utilización*** por parte de los usuarios finales en el mundo real (E11). La solución es puesta a funcionar en la esfera del usuario, comunidad o contexto de utilización para la cual fue concebida y retorna al sistema de investigación con

---

<sup>102</sup> Ver glosario

<sup>103</sup>“ La Cámara de Diputados estrena en su portal una herramienta para facilitar la transparencia y la interacción de la institución con la sociedad: el robot digital Ulises. Combina diversos algoritmos de inteligencia artificial, que permitirán cotejar automáticamente diversos datos legislativos disponibles en el portal de la Cámara y en diversos sistemas de la misma”

<https://www.camara.leg.br/noticias/548730-camara-lanca-ulysses-robo-digital-que-articula-dados-legislativos/>-accedido en septiembre 2023.

nueva información, conocimientos y demandas (ver el recorrido de la línea azul en el esquema de la figura 52).

Este proceso puede ser retroalimentado infinitamente ( $E_{11} > E_1 + E_2$ ), ya que es el uso de un instrumento (y como consecuencia la producción de más y nuevos datos) lo que confiere dinamicidad a la vez que genera nuevos problemas. Es decir, las soluciones son las responsables, en gran medida, por mantener dinámico el espacio-tema. La mantención de estos programas y del buen funcionamiento de los modelos y de los algoritmos a para mejorar y ajustarse a la realidad de su uso, requiere la retroalimentación de los nuevos datos y de un nuevo procesamiento de datos. Se le obliga, entonces, a una gestión y esfuerzo constante de mantención del sistema, sea desde el punto de vista financieros, como en términos de capacidad operacional y capacidad de gestión y gobernanza de datos.

Hasta el momento, todos los proyectos que acompañé, hablaban bajo el presupuesto del mantenimiento de las soluciones, incluso después de haber alcanzado los objetivos esperados. El proyecto investigativo siempre se termina, generalmente por cuestiones burocráticas relacionadas al financiamiento, pero para mantener la estructura de operación de la solución, se imprime la obligación de continuidad de mantención del sistema con cierta regularidad.

En cada ciclo, definido en función de las características de cada proyecto y su respectivo financiamiento, se busca un consultor (C), un experto técnico (ExTec), o un científico de datos (CD) para contratar por el proyecto, a fin que realice el calibramiento de los modelos algorítmicos generados, en base a los nuevos datos de entrada generados a partir, precisamente, de la utilización concreta de la solución, es preciso garantizar la infraestructura mínima necesaria para que la solución o producto desarrollado continúe “corriendo”.

Para garantizar la efectividad de un proyecto es necesario que todas las partes involucradas comprendan mínimamente el desarrollo y ejecución interna de la investigación, pues el procesamiento computacional de los datos resulta relativamente complejo para la población no letrada en estos temas. Es importante no solamente transparentar los procesos de investigación, los modelos elegidos, las decisiones tomadas, sino también asegurarse de que el usuario final sea capaz de entender los resultados, de visualizar la potencialidades de un proyecto y comprender la importancia de su implementación.

En el caso de proyectos vinculados al servicio público, lo más común es que los costos de la operación de la solución fueran incorporados a los gastos corrientes de la organización. Estos

gastos por lo general son distribuidos entre las siguientes ocupaciones: salarios del personal en TI, costos de licencias y adquisición de softwares asociados al desarrollo, pago de horas de procesamiento en línea o almacenamiento en nube, y por último, también con la adquisición de equipamiento necesario a la operación de la estructura de procesamiento y almacenamientos de la información. Todo esto cubierto bajo gastos operacionales del servicio en cuestión.

Esta operacionalización de los datos de gobierno en las investigación en ciencias de datos en contexto de su servicio público es tan vasto y profundo que fácilmente podría ser tema para otra investigación. Sin embargo, en el transcurso de las observaciones, abordaré con más refinamiento el tema de los grandes datos de gobierno involucrados en la investigación aplicada a ciencia de datos en América Latina.

### 8.3 Profesionales: los humanos por detrás del funcionamiento de la máquina laboratorial

A lo largo de la descripción del ambiente socio técnico de investigación social en ciencias de datos se presentaron los diferentes actores humanos participantes de dicho sistema, que sintetizo a continuación.

**Tabla 26.** *Actividades profesionales involucradas en el sistema Investigación social en Ciencia de Datos.*

Actividad	Descripción de la Actividad	Asociación en la Figura	Características
<b>Gestor de Proyecto (GP)</b>	Proponente del proyecto Detector de la base de datos de entrada Gestor de recursos y financiamientos proveniente del servicio que participa	Esfera Problema (E2)	Profesional vinculado a algún servicio público Generalmente Jefe o coordinador de una unidad de TI o innovación en la institución a que pertenece
<b>Académico Gestor (AcG)</b>	Investigador académico especializado en temas de sociales en ciencia de datos.	Esfera Problema (E2)	Aporta experiencia académica y metodológica a la investigación Dirección y movilización de recursos externos a la parte contratante Principal punto de movilización y contacto con desarrolladores, científicos de datos y desarrolladores

			Coordina actividades, asigna recursos, establece puentes entre distintos profesionales y garantiza que se cumplan los objetivos del proyecto.
<b>Experto Técnico (ExTec)</b>	Profesional con experiencia técnica en ciencia de datos que colabora en proyectos de investigación.	Esfera Problema (E2)	Ofrece conocimientos técnicos para la recopilación, procesamiento y análisis de datos. Informa sobre la infraestructura de operación y gestión mínima necesaria para la realización del proyecto
<b>Especialistas (Esp 5)</b>	Expertos en áreas específicas relevantes para el proyecto de investigación Especialistas en temas sociales y de otras áreas del conocimiento	Esfera Problema (E2)	Contribuyen con conocimiento disciplinario en sus áreas de especialización para contextualizar los problemas, datos y resultados. Fuerte Letramento tecnológico en el caso de los especialistas en Ciencias sociales
<b>Científico de Datos</b>	Profesional especializado en ciencia de datos Con experiencia en análisis de datos y desarrollo de modelos predictivos. No necesariamente utilizan este título profesional.	Esfera Datos (E1)	Utiliza herramientas, tecnologías y metodologías en ciencia de datos para analizar y extraer información de una determinada base de datos, identifica patrones y puede ser capaz de desarrollar modelos.
<b>Consultor</b>	Profesional que brinda asesoramiento y orientación puntual en proyectos de investigación en ciencia de datos. Puede ser un consultor de tipo técnico (involucrado con logística y operación computacional) o de tipo especialista (experto en el tema involucrado a investigación)	Esfera Problema (E2)	Proporciona perspectivas externas, asesoramiento estratégico y apoyo en la planificación y ejecución del proyecto.

Cada uno de estos roles colaboran en la investigación social en ciencia de datos, aportando sus conocimientos y habilidades específicas para abordar el problema en cuestión y generar resultados significativos a los proyectos de investigación como un todo.

Con respecto específicamente a los científicos de datos (CD), es importante dejar claro que este título representa más un cargo que una profesión. Esta ocupación puede ser ejecutada por distintos profesionales, por lo general, son ingenieros, la mayoría en sistemas o ingenierías de la computación pero también es posible encontrar proyectos donde el científico de dato define su ocupación como: Ingeniero de datos, analista de datos, Ingeniero matemático, o ingeniero de machine learning.

La responsabilidad de la calidad de los datos suele ser compartida entre los equipos de ciencia de datos y los demás profesionales y organizaciones contratantes. Cada uno de estos roles desempeña un papel importante en el ciclo de vida de los proyectos de investigación social y el procesamiento computacional de grandes datos.

#### **8.4 Un sistema socio-técnico: la manera GobLab de funcionar**

Una vez identificado el patrón de funcionamiento de los distintos proyectos conducidos al interior de GobLab descrito en la sesión anterior, fue evidente la semejanza de este esquema figurativo a los sistemas sociotécnicos descritos y conceptualizados por Latour (1992; 2008; 2013) en sus innumerables estudios sobre ciencia.

La observación de este patrón resulta de mi experiencia laboral en proyectos de investigación en ciencias de datos, y de las entrevistas realizadas a los científicos autores de publicaciones en big data de Chile, Brasil, Ecuador y México, gracias al análisis bibliométrico Scielo-Scopus realizado, y a mis visitas a diversos espacios de investigación semejantes en Brasil. Todo esto proporcionado como consecuencia de mi investigación doctoral.

Juntas, estas experiencias me involucraron en un escenario de investigación científica muy similar a lo que en los estudios constructivistas de la ciencia, se atribuye a los sistemas o redes sociotécnicas (Thomas y Buch, 2008; Latour, 2008). A pesar de mi declarada preferencia por la obra de Knor Cetina (2005), terminada la etapa de recolección de información y ya iniciado el proceso de análisis de los datos de la investigación, es casi imposible no traer a la discusión el hecho de haber estudiado, de haber estado presente observando un escenario sociotécnico en mi propio hallazgo de investigación.

Con todo, esta percepción quedó clara conforme construía el esquema (figura 52) en mi proceso de análisis de la información obtenida etnográficamente. Lo presenté en variadas instancias académicas, teniendo así la oportunidad de adaptarlos, siempre en diálogo con la retroalimentación de mis pares científicos CTS interesados en big data y temas afines.

Cuando presenté por primera vez la “versión 1.0” de este esquema al núcleo de investigación<sup>104</sup> del que además soy parte, al final de mi intervención, un compañero de la

---

<sup>104</sup> Núcleo Especulativo de las Tecnociencias y Biomateriales, coordinado por el Profesor Adrés Gomez Seguel, mi profesor guía y orientador de investigación doctoral. Para saber más: <https://sites.google.com/view/ctb-s/inicio?authuser=0>

audiencia me preguntó si ya había dado un nombre a este modelo. Complementó que ilustraba muy bien la dinámica de las redes sociotécnicas conceptualizadas por Latour (Latour, 2007; 2008) y los constructivistas. De hecho, gracias a esta intervención pensé por primera vez sobre esta relación.

Me fue necesario retomar, entonces, la lectura de Latour y sus estudios de ciencia y laboratorio, para reencontrarme con su perspectiva en cuanto a la forma de abordar y analizar tanto la información generada en campo, como los problemas propios a los estudios semejantes a lo que estaba emprendiendo. A partir de la relectura de estos clásicos textos latorianos (Latour 1992; 2008; 2011; 2013) me quedó más claro cómo se fue constituyendo esta investigación. Resultó ser un bonito ejemplo de como conceptos ya establecidos, cobran aún más sentido y vida a proyectos innovadores, aun en la contemporaneidad.

Textos como *Ciencia en acción* (1992) y *Reensamblar lo social una introducción a la teoría del actor-red* (2008), se constituyeron como herramientas teóricas muy importantes para el momento de análisis de la información generada en mi propio estudio. Orientándome, promueven también el estrecho diálogo entre la discusión que ya vengo levantando a través de la obra de Knorr-Cetina (2005) y todo el desdoblamiento que tenemos hechos con respecto a los estudios de la ciencia, tecnología y sociedad.

Es teniendo como base tal discusión que se puede originar dicho esquema: una interpretación de un espacio de producción sociotécnica del saber, siendo yo misma su testigo-fabricante. Es así entonces que más de 40 años después de los primeros estudios de laboratorio y la ciencia y producción del conocimiento científico y social, acá estoy yo dialogando con los mismos conceptos, patrones y agencias, estudiados por estos clásicos autores.

El primer aspecto a resaltar en cuanto a su adecuación conceptual a los sistemas sociotécnicos, es la interconexión de actores humanos y no humanos (Latour, 1992) que claramente ha estado presente en todo el tono de la investigación. Llegamos a abarcar esta temática en la primera parte del trabajo, en la discusión con respecto a la importancia de la agencia de los grandes datos (artefacto no humano) durante todo el proceso de investigación científica social pautada en el procesamiento computacional de los datos a partir de big data.

En una red sociotécnica (Latour, 2007), se considera actores humanos a los científicos, los usuarios, los especialistas, gestores de proyectos, los distintos funcionarios públicos trabajando para una agencia de un Estado latinoamericano, por citar ejemplos de los casos

estudiados acá. Los actores no humanos serían las tecnologías, los modelos, códigos y algoritmos, las computadoras, las diversas tablas de datos, artefactos y sistemas; los que están interconectados y son igualmente importantes en la configuración de la acción y los resultados (Latour, 1992). No solamente esto, también están dotadas de vida, de agencia, importancia y significado: “Miren alrededor: los objetos científicos circulan como sujetos, objetos y discursos a la vez. Las redes están llenas de seres. En cuanto a las máquinas, están cargadas de sujetos y de colectivos” (Latour, 2007, p. 101).

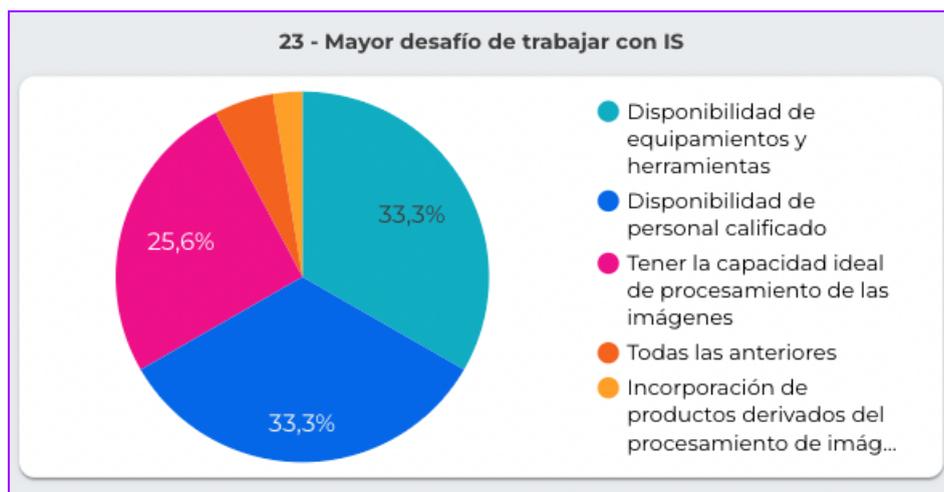
Hemos visto durante la descripción del sistema representado a través de la figura 52 que la interconexión entre la capacidad de infraestructura (Cp3) y la capacidad de gestión y gobernanza de los datos (Cp4) están impregnadas de actores humanos y no humanos, conectados a partir de una íntima relación de dependencia. La logística de datos no puede funcionar en separado de la estructura y la operación conlleva reunir los más distintos profesionales y expertos humanos para llevar a cabo los proyectos de investigación de esta naturaleza.

Un hallazgo que puedo contribuir para evidenciar tal relación es uno de los resultados de la encuesta a los usuarios de imágenes satelitales en Chile, derivada del proyecto SAMSARA, donde fui investigadora cualitativa<sup>105</sup>. Cuando preguntados sobre los mayores desafíos en cuanto al trabajo con las imágenes satelitales (un ejemplo de tipo de información big data que requiere procesamiento computacional) observamos el mismo grado de insatisfacción en cuanto a disponibilidades de equipamiento y herramientas (actores no humanos) y la disponibilidad de personal cualificado (actores humanos), como demuestra se ilustra a continuación.

---

<sup>105</sup>Para acceder a los datos de la encuesta y su dashboard interactivo, acceder a : [Encuesta Imagen Satelital SAMSARA](#). Para acceder al los resultados del estudio, a través del informe de investigación, acceder a : [Informe Final Inv.Calitativa\\_SAMSARA](#)

**Figura 55.** *Desafíos IS, Proyecto Fondef SAMSARA, 2022.*



*Nota: Resultado investigación SAMSARA*

Como se observa, incluso los problemas y desafíos a enfrentar por proyectos en ciencias de datos, llevan imbricado la íntima relación entre el mundo de las cosas y el de los sujetos. Esto se percibe en varias de las entrevistas a los científicos entrevistados: un proyecto de investigación en ciencias de datos no se sustenta sin la agencia y la múltiple afectación de sus más variados actores. Son igualmente esenciales para la mantención e incluso estabilidad de todo este sistema de investigación alrededor de los grandes datos.

El primer desafío con relación al funcionamiento ideal de nuestra unidad en ciencia de datos y IA del senado, no fue solamente tener que esperar y permitir que durante 2, 3 años nuestros profesionales contactados, concursados con más de 10 años de casa, terminaran de hacer sus especializaciones y posgrados. Mientras tanto, trabajábamos internamente para sensibilizar a los gestores, o nuestros superiores directos, de la importancia de aprobar financiamiento destinada a la compra de equipamientos. De nada servía que mi equipo volverá máximamente capacitado si en términos de estructuras estábamos como en la edad de la piedra. Estas dos cosas tienen que caminar juntas: capacidad humana y capacidad de estructura de equipamiento, de computadores, de softwares y logística de seguridad de datos por supuesto. (Conversación, Visita Ulysses, Brasil, 2022 - Mi traducción)

Por otro lado, la interpretación de la figura 55 también puede sugerir otro aspecto importante relacionado a la forma laturiana de concebir las redes sociotécnicas: la igualdad de condiciones entre los actores. En el análisis de las redes sociotécnicas, no se delega prioridad

a los actores humanos sobre los no humanos (Latour, 1992). Todos los actores se consideran en igualdad de condiciones en la configuración de la red, aunque a partir de un estudio más profundo se revele la contrariedad de este escenario.

La importancia simétrica entre hombres y máquinas es parte de los resultados de la categorización del sistema que vengo estudiando. La recién transcrita entrevista da cuenta de este aspecto: es posible percibir que la impronta de un proyecto en ciencias de datos debe considerar mantención, estabilidad y desarrollo en esferas tangibles tanto a los humanos como a los no humanos. Están en igualdad de importancia, de influencia e inclusive, en el mundo de los resultados (7).

Un científico más experto es capaz de hacer mejores elecciones de metodología, es más sensible y expedito en relación a la punta del sistema y los usuarios, ya sabe comunicar mejor los desdoblamientos del procesamiento y de los resultados de la información generada. Paralelamente, cuanto mejor y más ajustada es la infraestructura computacional, de equipamiento, de datos y gestión de sistemas, se esperan informáticos de mejor calidad. Son interdependientes e iguales en importancia.

Hombres y máquinas, en esta realidad tienden a ser peculiarmente especializados: la totalidad de los científicos estudiados, y con los cuales pudimos tener contacto a través de la investigación, son posgraduados. La gran mayoría poseen título doctoral y acaban especializándose en dimensiones muy particulares de sus disciplinas, como suele pasar naturalmente con cualquier académico altamente titularizado. Por otro lado, las máquinas y sistemas de infraestructura de operación no humana son igualmente específicas, costosas y muchas veces “hechas a la medida”. Es necesario que sean compatibles, disponibles y orquestadas de la manera más funcional posible, lo que requiere también un alto grado de especificidad en los equipamientos y proceso de montaje, compra y contratación de sus servicios.

El segundo punto donde veo equivalencia con las redes sociotécnicas alrededor de la construcción del conocimiento científico trabajadas por Latour, es el lugar de la contextualización, o condición situada (Knorr-Cetina, 2005) de las conformaciones de estas redes. La descripción de una red sociotécnica latouriana nos enseña sobre cómo son altamente contextuales y varían según el tiempo histórico y el espacio contextual en donde están insertadas. Lo que funciona en una red sociotécnica en específico, puede no funcionar de la

misma manera en otra, y esto depende de factores sociales, culturales y políticos (Latour, 2007).

Aunque en los diferentes proyectos de investigación en ciencias de datos que pude acompañar me demostraron cierta regularidad de comportamiento, aunque trabajaran con diferentes datos y objetivos, están regidos bajo un paraguas contextual muy evidente y cambiante: Si cada proyecto es conducido, priorizado y diseñado a partir de la base de datos que se tiene en manos, es más probable que cada uno de estos sistemas sea único, ya que sus datos originales están muy asociados a un problema específico directamente relacionado a la manera de capturar la realidad desde donde provienen dichos datos.

De esta realidad, deriva una tercera característica compartida, su heterogeneidad: las redes sociotécnicas pueden ser muy heterogéneas, ya que además de involucrar una amplia variedad de actores, también involucra a otras dimensiones asociadas, como, por ejemplo, las diferentes organizaciones en el proyecto, tecnologías, normas, regulaciones y más. La heterogeneidad de las redes socio técnica estudiadas, como pudimos percibir, no involucran solamente las peculiaridades propias del contexto social, histórico y político involucradas en cualquier actividad científica, sino que también posee un grado aún más fuerte de contextualidad debido a especificidades únicas que conllevan la esfera de datos y todos sus elementos orbitantes (E1) vinculados y delimitados a partir de su relación con el problema de la investigación (E2).

Respecto a la heterogeneidad, es importante dilucidar el alto grado de convivencia de las distintas organizaciones al interior de los proyectos de investigación en ciencia de datos. Esto ocurre no solamente por la naturaleza de su estudio aplicado, lo que naturalmente obliga a una convivencia institucional abierta a estos proyectos de investigación, también porque siendo proyectos interdisciplinarios, que involucran diversas expertises alrededor de los especialistas y científicos involucrados, genera un ambiente institucional diverso, heterogéneo e interdisciplinario. Hablaremos más específicamente sobre esta dimensión en las siguientes secciones, donde discutiremos los desdoblamientos metodológicos de investigar aspectos sociales a partir de las ciencias de datos.

Por fin, y no menos evidente, está el hecho de que a partir de la mirada de las redes socio técnicas se puede entender con aún más profundidad el proceso de co-construcción entre tecnología y sociedad. La definición de la no neutralidad de las herramientas tecnológicas (Latour, 1992) es el panorama teórico ideal para enmarcar el sistema representado por esta

investigación. El concepto resalta que las tecnologías no siendo neutrales, moldean y dan forma a la sociedad al mismo tiempo que la sociedad les da forma. La tecnología y la sociedad evolucionan en una relación mutuamente constitutiva y podemos así entonces, expandir el raciocinio para la mutua construcción de los saberes científicos y las estructuras sociales.

Discutimos en capítulos anteriores (Cap. 8 - sección 8.2) porqué los laboratorios y los estudios de ciencia se conforman como un poderoso sistema sociotécnico de interpretación de la sociedad. Conferimos, bajo los resultados de la investigación conducida, que sociedad y laboratorio son reflejo del mundo real y de las demandas sociales contemporáneas. Partiendo de la necesidad de entender las tecnologías, los estudios de ciencia y tecnología como parte de un proceso histórico social único, depende del contexto y condición situada de la investigación.

En base a todo lo dicho anteriormente, definiendo esta investigación y el sistema esquemáticamente presentado y fabricado como un interesante estudio de caso con respecto a los sistema sociotécnicos de Latour (1992; 2008; 2007; 2013). Como diría el autor, los estudios de caso son una manera de aplicar el concepto de red sociotécnica, y es esto lo que está siendo presentado en la parte 2 de esta tesis: investigando y analizando cómo se forman y evolucionan estas redes en situaciones concretas, como el desarrollo de tecnologías, la implementación de políticas o la resolución de problemas tecnológicos, viene haciendo parte de todo este esfuerzo investigativo.

Poder enmarcar la representación presentada bajo el concepto de red sociotécnica, en base a mi estudio, ha sido especialmente importante para entender la dinámica de la innovación tecnológica, la toma de decisiones, la política tecnológica y la adopción de tecnologías en la sociedad. Ayuda a desafiar la noción de que la tecnología es independiente de la sociedad y subraya la interdependencia entre ambas, lo que por sí mismo ya sería un valioso aporte a los estudios de ciencia, tecnología y sociedad contemporánea.

### **8.5 Qué puede enseñarnos un laboratorio social en ciencia de datos: Implicaciones metodológicas a la investigación social con vocación data science**

Diferentes prácticas y dinámicas de investigación configuran distintas maneras de producir conocimiento. La introducción de las técnicas de ciencia de datos, big data y de procesamiento computacional de datos digitales, ha cambiado la forma como se obtiene y se produce información, y, como hemos visto, también está cambiando la manera en que se hace ciencia (Gualda, 2022). El sistema-sociotécnico que se configura alrededor del big data incentiva la convivencia entre distintas disciplinas, especialidades y lenguajes. Estos coexisten en un entorno material, compuesto de máquinas, sensores, nubes y bases de datos que interactúan constantemente con investigadores, ingenieros, gestores, diseñadores, tomadores de decisiones y otros actores humanos.

Los resultados presentados en esta investigación, tanto sobre la producción científica, como sobre las prácticas de investigación en ciencias sociales en Latinoamérica, sugieren que estamos frente a un proceso de transformación de la interacción entre los actores (humanos y no humanos) que producen ciencia.

Una vez que se entiende bien la dinámica investigativa al interior de este sistema socio técnico de investigación en ciencias de datos, es más fácil percibir que diferentes prácticas y dinámicas de investigación implican distintas maneras de fabricar el conocimiento. Desde esta base, se hace lógico establecer una relación entre la artesanía de la investigación y las técnicas computacionales de procesamiento de datos, con la formas cómo se obtiene y se produce información útil y científica. Los cambios de los nuevos escenarios, o de nuevos sistemas al interior de la investigación, acaban por producir y retroalimentar cambios en la sociedad, la toma de decisiones, y en la gestión de la vida; así como también impactar la forma en cómo se está haciendo ciencia al interior de los laboratorios.

El entorno investigativo en ciencia de datos empuja, como hemos visto, la convivencia entre distintas especialidades, lenguajes y disciplinas del saber. Coexisten en un entorno material, lleno de máquinas, cables, equipamientos y capacidad humana de un pensar y trabajar tecnológico. Así mismo, observamos que en el mundo de la ciencia existe una forma peculiar de existir y hacerse presente frente a la digitalización y desarrollo de la vida social. En el universo científico, su interferencia, entonces, no podía dejar de ser evidente.

Bajo la prerrogativa de que la introducción de nuevas técnicas, saberes y disciplinas al interior de la investigación influencia, dando forma y contorno a las prácticas de producción de información, encontramos que la introducción del ecosistema big data a las investigaciones en ciencias sociales, tiene algunas implicaciones metodológicas en nuestra manera de hacer ciencia. Los resultados presentados en esta investigación, tanto sobre la producción científica, como sobre las prácticas de investigación en ciencia sociales en

### ***8.5.1 Focalización en la investigación aplicada.***

Los resultados evidencian que la investigación social producida a partir de la ciencia de datos tiende a enfocarse más en la resolución de problemas prácticos que en la adquisición de conocimientos teóricos o abstractos, teniendo como objetivo principal resolver problemas prácticos (principio de la incidencia). En lo siguiente fragmento de discurso expresa muy bien esta primera implicación:

“Su trayectoria [refiriéndose al decano de la UAI en 2016] siempre ha sido de promover la modernización del Estado. Él quería crear en la escuela, un centro que se dedicara a la modernización del Estado, no a la publicación académica, sino a lo que en la universidad se conoce como vinculación con el medio”. (MPH, Entrevista 32/2022 - Chile. mi traducción)

Al enfocarse en la resolución de problemas prácticos, la investigación en ciencia de datos se aleja del desarrollo de teorías abstractas y se enfoca en incidir en algún problema real implicado en la estructura que los datos disponibles proporcionan. En este sentido, la tarea de definir claramente el problema a resolver se vuelve tan importante como la de capturar, uniformizar, limpiar y manipular las bases de datos (es decir, la calidad de los datos).

El interés por profundizar en un tema y conjeturar sobre el asunto que se desea conocer, no está en la prioridad de los trabajos de investigación aplicada en ciencia de datos. El tema de la investigación y las teorías por detrás implicadas, son más bien una especie de paraguas que abriga y hace convivir los datos, los problemas y la definición de los diferentes especialistas involucrados en el proyecto. Entendemos, por lo tanto, que el problema/objetivo de investigación, en la gran mayoría de los casos, está orientado a la incidencia y resolución de problemas a partir del conjunto de datos que se tiene o que se puede acceder. Todo en la investigación apunta hacia esto.

La investigación aplicada permite a los científicos utilizar los datos disponibles para identificar y abordar problemas específicos que afectan a la sociedad, como mejorar resultados, evitar fraudes, distribuir equitativamente los recursos, prever los gastos energéticos del futuro, apoyar el defensor público en sus labores y crear alertas y sistemas de monitoreo ambiental, por ejemplo. Por lo anterior, el acceso a los datos es fundamental, pudiendo provenir de diversos campos y fuentes (medicina, industria, medio ambiente, etc.), y ser extraídos con diferentes tipos de obtención de datos (estáticos, dinámicos, en tiempo real).

Es interesante notar que tal tendencia enfocada a lo práctico no acaba por expresarse en las publicaciones académicas del área. Como puede verse en el estudio bibliométrico en Scielo y Scopus, aunque las investigaciones sociales en/para/con la ciencia de datos posea este carácter pragmático y aplicado, la tendencia no se traspasa a las publicaciones. Aunque los científicos sociales están interactuando en un sistema completamente focalizado a los resultados, práctica, incidencia y diversidad metodológica, sus publicaciones siguen teniendo una marca teórica muy fuerte. A la hora de publicar, el profesional de ciencias sociales todavía opta por escribir textos que problematizan conceptos e implicaciones teóricas y epistemológicas vinculadas al contexto de investigación en ciencia de datos.

La explicación para tal puede encontrarse en dos explicaciones. La primera, es la dificultad de encontrar un área del conocimiento académico ya institucionalizada capaz de dialogar con este tipo de saber. Así lo relata una científica social entrevistada, doctora en ciencias políticas e integrante de diversos proyectos en procesamiento de lenguaje natural en grandes corpus textuales:

“El problema de estar también cómo estoy es que nadie sabe bien qué hago. Uno de los problemas de la interdisciplina, es, justamente, la parte de que te cuesta un mundo publicar, haciendo que también sea raro hacer contactos, hay que ser alguien muy específico y que no está trabajando en otras cosas y no tiene tanto tiempo... en mi caso realmente es un desafío publicar.” (MPR - Entrevistas 15/2020 - Chile)

La segunda explicación está relacionada con que hay una falta de incentivo de parte del medio como también hemos visto, las publicaciones en formato académico (artículos) no son prioridad en estos laboratorios.

Ellos [los distintos especialistas asociados a los proyectos del laboratorio] trabajan en temas que a nosotros nos interesan y nosotros en el fondo los

apoyamos en llevar a cabo sus proyectos. Pero a cambio ellos hacen cosas también de políticas públicas, y no solamente sus *papers* académicos.

**Entrevistador:** ¿Ustedes como laboratorio no publican *papers* académicos? No es su interés principal?

**Respuesta:** No es nuestro interés principal, pero tenemos, en general, sólo publicamos modelito así académico cuando alguno de ellos lidera la escritura porque para ellos como académicos es importante”. (MPH, Entrevista 22/2022 - Chile. mi traducción)

Así, la publicación es casi siempre una iniciativa proveniente de los científicos vinculados al proyecto en cuestión, pues para ellos las publicaciones son una oportunidad de difundir un aspecto puntual de sus propias experiencias o de alguna inspiración generada en el contexto investigativo del laboratorio. El interés va orientado a las elecciones individuales de los diferentes actores, y no como una exigencia de la estructura de funcionamiento y mantención del laboratorio, que por su parte, está mucho más interesado en las publicaciones ejecutivas: guías de implementación, protocolos de procedimientos, cartillas de recomendaciones, informes de los proyectos ejecutados.

Defiendo que el enfoque pragmático generado de tales condicionantes están cambiando la manera en que se abordan los problemas de investigación. En lugar de simplemente identificarlos y analizarlos, los investigadores se ven interpelados a actuar y resolverlos de manera más eficiente y con mayor capacidad de expandir las técnicas de obtención de información relevante. Esta posibilidad resulta muy atractiva para muchos de los científicos entrevistados. En general, después de tener el primer contacto con el procesamiento computacional de los datos, valoran la potencialidad de introducir tecnologías similares en sus propias investigaciones. Entonces, se motivan a buscar maneras de adaptar los métodos utilizados para producir el conocimiento adquirido, y establecen contacto con las redes construidas durante la ejecución del proyecto como estrategia para armar nuevos proyectos de investigación, más afines a las temáticas que les interesan.

“Nosotros llevamos 2 años para fabricar e implementar este sistema que te estoy mostrando [Apuntando a la pantalla del computador con el sistema bem-te-vi abierto]. Hubo mucha gente involucrada. Llegó un punto que tuvimos que pedir asesoría a un grupo de investigadores matemáticos de la UnB [Universidad de Brasilia] para armar el modelo que necesitábamos para que el sistema funcionara bien, ya que lo que queríamos lograr superaba nuestras “capacidades intelectuales”. Resulta que al final salió todo tan

redondo y limpio que estos mismos científicos empezaron a utilizar los algoritmos creados en sus otras cosas, sus investigaciones... Lo que sé es que ellos en más de una oportunidad nos decían que la mejor solución para el problema que ellos estudiaban en sus laboratorios, fue generada a través de las posibilidades presentadas a partir de nuestra data y las soluciones que buscábamos en este momento. Y ahora, con algunas adaptaciones, por supuesto, implementaron los modelos de nuestra solución a algunos de sus trabajos. ¡Parece que nosotros les cambiamos la vida por allá!” (CMB - Entrevista 57/2023 - Tribunal Superior del Trabajo - Brasil - mi traducción)

De esta implicación deriva también el **principio de la incidencia**, es decir, el imperativo de la resolución de problemas: no saber, actuar. Según mi comprensión, el uso de ciencia de datos en investigaciones sociales aplicadas ha llevado a una nueva forma de enfocar los problemas. En lugar de simplemente identificarlo, analizarlo para conocer en profundidad un fenómeno social o una dimensión de la sociedad, los investigadores se enfrentan con la urgencia de interpretar un fenómeno a partir de sus datos con el fin de solucionar o mejorar dicho fenómeno. En este sentido, la metodología ha ido cambiando para que los investigadores trabajen de manera más práctica y orientada a la fabricación de “soluciones” o “productos” y no necesariamente a la fabricación de conocimiento.

En resumen, el enfoque de la ciencia de datos se centra en la resolución de problemas y la aplicación de metodologías prácticas, dejando en un segundo plano las discusiones teóricas y epistemológicas subyacentes a dichos problemas. La importancia de la resolución de problemas para proyectos de investigación en ciencia de datos se vuelve crucial, y se adopta el lema "actuar para resolver". Pasamos de las preguntas *¿qué puedo conocer más?* o *¿cómo puedo profundizar en el entendimiento de esta dinámica social?* a *¿qué tipos de problemas podemos resolver con lo que ya sabemos o podemos saber?* y *¿qué crees que se puede hacer con la información que tenemos?* La lógica de estudio, entonces, prioriza la solución de problemas, la base de datos se vuelve crucial, y prevalece una estructura no jerárquica entre teoría y metodología.

### ***8.5.2 Diseño de investigación limitada a la disponibilidad de los datos.***

El creciente uso de tecnologías de procesamiento computacional de grandes datos en el campo de las ciencias sociales, redefine la forma en la que tradicionalmente se ha abordado la investigación en ciencia deductiva (Popper, 1934). En lugar de establecer preguntas de investigación de manera independiente, se plantea un diseño investigativo que se ajusta a la

disponibilidad de datos. Esta evolución implica una reorientación hacia la resolución de cuestiones prácticas basadas en la riqueza de datos disponibles.

Los datos, en este contexto, desempeñan un papel central al determinar los problemas específicos que pueden ser abordados y las preguntas que pueden ser respondidas. En otras palabras, son los datos los que moldean la dirección de la investigación. La perspectiva guiada por el dato se basa en la idea de que los datos, cuando se analizan de manera adecuada, pueden revelar patrones, tendencias y relaciones significativas que de otro modo podrían pasar desapercibidos.

Esta transición hacia un diseño investigativo basado en datos no sólo exige una consideración más profunda de la fuente y la calidad de los datos disponibles, sino también la habilidad para formular preguntas relevantes que aprovechen al máximo el potencial informativo de esos datos. En última instancia, la sinergia entre la disponibilidad de datos y la capacidad de plantear preguntas significativas abre nuevas oportunidades para la generación de conocimiento en las ciencias sociales, permitiendo un enfoque más adaptado y aplicado a los desafíos contemporáneos.

Este entendimiento viene siendo cada vez más consensuado entre los especialistas, sobre todo los de ciencias sociales que trabajan en proyectos de investigación alrededor de tecnologías big data. En el siguiente extracto se resume la opinión de algunos de los científicos entrevistados:

Ahora, desde el punto de vista metodológico, concretamente el big data, nos permite hacer una recogida de datos exhaustiva y sólo pensar en cómo vamos a analizarlos después de la recogida. Algo que antes no era posible, pues ya tenías que preparar tu investigación, parametrizar todo lo que ibas a llevar al campo para analizarlo después, antes de ir al campo. También hay un trabajo previo a la recolección con big data: obviamente no vas a apretar un botón y todo te va a caer del cielo. Pero una vez recogidos, las posibilidades de análisis se vuelven infinitas en cuanto a lo que se puede hacer, mientras que los análisis con datos recogidos de la forma tradicional ya están limitados por lo que el investigador se propuso hacer al principio de la parte epistemológica de la investigación. (PLC - Entrevista 09/2023 - Brasil - mi traducción).

Esta manera de conducir los proyectos de estudio difiere de la investigación clásica, la cual plantea una hipótesis a partir de un marco teórico, para luego comprobarla con datos que genera posterior a su pregunta. Justamente, la primera diferencia significativa entre estas dos maneras de entender la construcción del conocimiento científico, está relacionada al origen

de las preguntas de investigación, pues, en contraste a la investigación clásica, en el modelo investigativo basado en datos, la pregunta de investigación surge, con frecuencia, directamente de los datos disponibles. Es decir, los investigadores primero explorarán sus datos para descubrir patrones y tendencias, antes de formular cualquier pregunta.

Lo mismo acontece con el levantamiento de hipótesis y la confirmación de teorías. Mientras que en el enfoque clásico, solemos formular una hipótesis derivada de una o más teorías, para luego buscar los datos que la confirman o refutan, en el modelo basado en datos, las hipótesis se desarrollarán en una etapa posterior a este proceso, ya habiendo analizado los datos. Esto puede permitir un enfoque de producción del conocimiento más inductivo (Hume, 2002), donde las teorías emergen de los datos en lugar de dictar la dirección de la investigación.

En este abordaje más inductivo de hacer ciencia, se percibe una mayor flexibilidad con respecto a la exploración y el descubrimiento de patrones inesperados en los datos. Los investigadores usan variadas técnicas de análisis y visualización de datos, permitiéndoles dialogar de una forma diferente con los resultados. Es a partir de este movimiento que empiezan a aparecer nuevas relaciones y fenómenos que de otro modo no habrían sido evidenciados, permitiendo así un ambiente propicio a la innovación.

Gracias a esta flexibilidad, junto al espíritu de experimentación que conlleva el modelo basado en datos, los métodos están siempre a prueba a través de un constante testeo metodológico de los sistemas, herramientas, modelos computacionales y algoritmos, lo que ayuda a su constante desarrollo.

Al respecto, leamos el siguiente comentario de una profesional en ciencia de datos entrevistada durante la trayectoria etnográfica. Ante la pregunta sobre cuáles son las mínimas condiciones necesarias para validación de un resultado desde el punto de vista del científico de datos, contesta:

Lo que yo hago, en el fondo, es admitir que mi propia idea, mi propio pensamiento juega un papel importante en la formulación de la hipótesis. Pero en la parte del procesamiento de datos, sé que la metodología tiene que ser lo más robusta posible. Lo más robusta posible en estadística significativa, por ejemplo, garantizar que haga pruebas y que vaya a ocupar muchos modelos alternativos. Que vas a ocupar variaciones de variables, variaciones de modelos, variaciones de sensibilidad, como para convencerte a ti misma y a los revisores el porqué está bien lo que está yendo bien! (MPR - Entrevista 36/2022 - Chile).

En frases como esta es que se corrobora que realizar pruebas o calibración de los modelos, es una actividad fundamental parte del procesamiento de datos, pues así su transformación en información es válida y consistente. A través de la implementación de las diferentes metodologías de procesamiento de datos existentes, el científico de datos, junto al equipo de especialistas, pueden ajustar más efectivamente los medios para producir y analizar la información, de acuerdo con la característica intrínseca a los datos, los resultados y las soluciones que se espera obtener.

En otra conversación con un científico de datos responsable de uno de los proyectos que acompañé, al preguntarle sobre la rutina de sus actividades en relación a las diferentes fases del proyecto, me explicó lo siguiente:

La primera pregunta es ¿qué modelo elijo? Existen muchos, y ahí es donde se origina el súper *vector machine*, la red neuronal que es un tipo de modelo, el *détection tree* es otro tipo de modelo, el *random forest*... Ya: ¿Cómo elegir el modelo, entonces?, depende de la actualidad y la aplicación. Por ejemplo, si los datos son un archivo en Excel, básico, es por experiencia del científico de datos de lo que ya ha trabajado, no elige mucho el *support vector machine*, el *random forest*, el *détection tree*, para esto. Cuando vas a hacer, por ejemplo, procesamiento de imágenes, procesamiento de lenguaje natural y cosas más avanzadas, por lo general las redes neuronales funcionan muy bien. Porque el modelo, para no explicarte la matemática, viene adaptado para eso, o sea, hay algunas redes neuronales convolucionales, hay redes neuronales recurrentes que emulan el comportamiento de la visión humana, emulan lo que es el habla, tienen memoria, y los demás modelos son como para tus estructuras. En resumidas cuentas, en este tipo de modelos, uno tiene una palabra última, ¡ni de Dios! Te toca probar; el cientista de datos tiene que empezar a probar, conocer con su expertiz. ¿Estamos claros hasta ahí? (RNN - Observación participante grabada - 2022 - Chile)

Esto también significa que los métodos de análisis de datos deben ser actualizados y mejorados continuamente para obtener resultados precisos y más ajustados a los nuevos datos (generados), lo que es una característica del funcionamiento de este sistema, como vimos en la sección anterior. A esto se suma la validación cruzada de los resultados utilizando diferentes técnicas y fuentes de datos, haciendo que la importancia de probar diferentes métodos y no aferrarse a uno sólo durante toda la investigación, sea esencial al proceso de adaptación a los distintos tipos de datos disponibles y al problema en cuestión.

Una tercera implicación resultante del proceso investigativo social en ciencia de datos es la implementación de proyectos que a menudo se asocian al abordaje de problemas contemporáneos y dinámicos, ya que poseen el potencial de aprovechar tecnologías big data y análisis computacional avanzado para procesar grandes y diversos conjuntos de datos. Asociado a esto está el hecho de poder utilizar datos recopilados en tiempo real o en intervalos cortos para comprender, posibilitando así maior agilidad y capacidad de incorporación de nubes y actuales datos, con la finalidad de responder a desafíos actuales de la sociedad, o comunidad adyacente a esta información.

A continuación, presento las principales diferencias encontradas entre la manera clásica de hacer ciencias sociales y la manera data driven de conducir un proceso de producción de conocimiento, en el contexto de observación de esta investigación:

**Tabla 27.** *Diferencias entre el modelo clásico y el modelo basado en datos*

Característica	Modelo Clásico de Investigación (Deductiva)	Modelo de Investigación Basado en Datos (Inductiva)
<b>Origen de Preguntas</b>	Derivadas de teorías y marcos teóricos sólidos	Derivadas de la disponibilidad de datos
<b>Hipótesis y Confirmación de Teorías</b>	Formulación de hipótesis basadas en teorías y búsqueda de datos para confirmar/refutar	Desarrollo de hipótesis después de explorar los datos
<b>Exploración y Descubrimiento</b>	Menos énfasis en la exploración y el descubrimiento de patrones inesperados	Mayor flexibilidad para explorar y descubrir patrones en los datos
<b>Uso de Tecnologías</b>	Menos dependencia de tecnologías big data y análisis computacional	Uso de tecnologías big data y análisis computacional avanzado
<b>Relevancia para Problemas Actuales</b>	Enfoque histórico y menor atención en los problemas contemporáneos	Enfoque contextual y mayor atención hacia los problemas contemporáneos
<b>Calidad y Accesibilidad de Datos</b>	Enfoque en la calidad de los datos, pero menos énfasis en la accesibilidad	Énfasis en la calidad, integridad y accesibilidad y reproducción de los datos

Ahora bien, si la teoría ocupa un lugar secundario en los proyectos de investigación social a partir de las ciencias de datos, pues, como ya vimos, se enfocan en resolver o dar soluciones efectivas a los problemas prácticos que los datos disponibles ofrecen., para qué sirve la teoría. Según los hallazgos, está al servicio de los resultados. No instruye sobre el problema a investigar, sino que forma parte del proceso de validación de los resultados, limitándose a:

i) Guiar la selección de variables y las técnicas de análisis de datos: los científicos de datos deben tener un conocimiento de las teorías relevantes para su campo de estudio, buscando identificar las variables de inclusión en el análisis de datos;

ii) Desarrollar hipótesis compatibles a las relaciones entre variables: casi siempre con la finalidad de predecir fenómenos futuros, teniendo como base el comportamiento del conjunto de datos disponible y trabajo, e

iii) Interpretar los resultados del análisis de datos. Al tener un conocimiento de la teoría y los fundamentos teóricos relevantes, los científicos de datos, junto a los especialistas, garantizan la validez y significancia de los resultados. Su adecuación teórica se construye como parte del proceso de interpretación de ellos.

Así, en resumen, la utilización de tecnologías en ciencia de datos en la imprenta científica implica diseñar la investigación a partir de los datos disponibles y las preguntas que pueden ser resueltas a partir de este conjunto de datos diversos. Los datos se utilizan para identificar los problemas específicos, mientras que las preguntas a responder demandas, necesidades o soluciones tecnológicas. El marco teórico utilizado es secundario y su función es en pos de obtener los mejores resultados prácticos posibles.

### ***8.5.3 Colaboración e interdisciplinarietà***

En este tipo de investigaciones, la metodología cobra más importancia que la teoría. A menudo, los investigadores invierten más tiempo en encontrar el mejor arreglo metodológico que les permita probar sus resultados, que en encontrar una teoría específica que explique los fenómenos analizados. Dado que no hay un método único de análisis y que la ciencia de datos tiene una amplia gama de técnicas metodológicas involucradas, existen diversas fuentes de modelos y formatos de datos que trabajar, haciendo que la colaboración y la interdisciplinarietà sean fundamentales en este ambiente de investigación. Cada disciplina aporta con diferentes perspectivas y formas de producir información, con la finalidad de

abordar los problemas, que, además, suelen cambiar su delimitación debido a la necesidad de ajustarse al encuentro de soluciones efectivas para la esfera práctica. En ese sentido, la colaboración entre investigadores de diferentes disciplinas ayuda a que los resultados sean significativos y relevantes.

La colaboración efectiva (casi obligatoria) entre los diferentes especialistas y profesionales contenidos en este sistema de investigación, ha creado un ambiente de profunda interdisciplinaridad<sup>106</sup> (Acosta, 2016), sobre todo entre especialistas y especialistas y científicos de datos. Los científicos de datos, o profesionales que los manipulan, no trabajan sin que les abastezcan, justamente, de datos. La elección de las variables a considerar en el modelo para *hacer la magia acontecer (rodar el modelo)*, en palabras de los nativos, también son dependientes de esta relación. Como demuestra el siguiente extracto de diálogo.

**Entrevistadora:** Bueno, es que estamos buscando entender ¿cómo saben si el modelo está bien? o sea, es solamente el especialista el que determina esto?

**Respuesta:** Sí, normalmente eso lo hace un especialista, o sea, normalmente necesitas el especialista, a excepción de que tengas *data* cruda, que no se haga nada con ella y hagas aprendizaje no supervisado, sacas *clustering*... pero, al final, necesitas, otra vez, otro experto para que te diga “oye, ¿y esta nubecita de datos qué me está diciendo?, ¿usted qué puede concluir de ahí?”. ¿Sí me entiende?

**Entrevistadora:** Entonces, en teoría, un cientista de datos nunca puede trabajar solo.

**Respuesta:** Pues sí, la verdad, nosotros nunca trabajamos, o sea, para proyectos grandes y serios, no trabajamos. (RNN - conversación informal - 2022 - Chile)

Por su parte, los especialistas, sobre todo los científicos sociales, son completamente co-dependientes de las capacidades computacionales del cientista de datos en cuestión para procesar la información brindada. Son también los especialistas los que, por lo general, toman decisiones sobre qué criterios, dimensiones, variables trabajar. La fiabilidad de un resultado

---

<sup>106</sup> En este enfoque, las disciplinas colaboran de manera más estrecha, buscando una integración de sus enfoques y métodos. Se busca un entendimiento más profundo del problema abordado al cruzar las fronteras entre disciplinas. Los investigadores pueden trabajar juntos para desarrollar un enfoque más holístico y complementario, utilizando las fortalezas de cada disciplina (Acosta, J., 2016).

proveniente del procesamiento de datos está notoriamente influenciado por la relación de las variables seleccionadas y su explicabilidad de acuerdo al funcionamiento “real del mundo”.<sup>107</sup>

[...] Entonces tú, como experta juegas en la entrada del modelo. Luego, en la salida, ¿cómo juega el experto? “Oye, me dijo 1.7, ¿para ti está bien?”. Tú, Bárbara, como experta dices “sí, el 1.7 está bien para ese fenómeno”.

**Entrevistadora:** Pero ¿cómo yo voy a saber que 1.7 está bien para ese modelo?

**Entrevistado:** No, pues, por lo menos digamos que, pensémoslo nuevamente el ejemplo anterior de las gallinas y huevos: es que por eso eres experta y tienes que conocer muy bien el problema. Si yo te digo “oye, yo tengo 2 gallinas de peso 4 kg y temperatura 30°C. Eso me dio una predicción de 5 huevos”. Entonces una es que tú tengas datos donde esto se te pueda validar. O sea, cuando tienes un algoritmo de inteligencia artificial, mira lo que uno llama “conjunto de entrenamiento”, con esto se entrena.

Y luego que el modelo está entrenado, tú le pasas estas variables y haces la predicción. Y él [el procesamiento, el entrenamiento del modelo] te da 31, te da 32, te da 34 y te da 37, y esta es la predicción. Entonces te digo: “oye, esto es lo que predijo y esto otro es lo real”, entonces tú dices “va bien. Está muy cerca” porque está...

Entonces ese “5 huevos” que salió acá [de la predicción] está bien para ti o porque tú de antemano conoces este resultado, lo tienes en otro archivo y lo conoces, o, es porque realmente eres experta en el fenómeno y tú dices “¡oye, sí! Yo he trabajado toda mi vida con eso y sí, más o menos con 2 gallinas de 4 kg y 30 °C, y efectivamente me producen esos 5 huevos. ¿Me entiendes? Esto lo defines tú [el especialista]. (RNN - Entrevista 41/2022 - Chile).

Es, entonces, el especialista quien tiene la capacidad de decidir la validez y fiabilidad de los resultados en base a su experiencia y estudio de la realidad alrededor del problema estudiado. Si bien el especialista juega un rol importante en relación a la validación de los resultados y fabricación de las soluciones creadas, la confección de los resultados, productos y soluciones provenientes de la investigación, también dependen de la capacidad del cientista de datos de entender las necesidades del proyecto: seleccionar el mejor conjunto de datos posible a ser

---

<sup>107</sup> Es importante esclarecer que la noción de realidad, en el ambiente de ciencia de datos, tiene mucho menos que ver con la verdad, y más con el sentido de aprehensión a partir de la experiencia. Como apunta Knorr-Cetina: “Aceptemos también que esas actividades de atribución de sentido tienen más en común con la ‘comprensión’, como un acto en el cual se funden la experiencia y la aprehensión teórica, que con la ‘explicación’, como la ‘aplicación de proposiciones teóricas a hechos que están establecidos independientemente mediante la observación sistemática’. También estos laboratorios, la experiencia y la comprobación de la ocurrencia de fenómenos patronizados en el contexto de observación es que es el indicador de relevancia y validez de los datos.” (2005, p.314)

procesado y comunicar los hallazgos encontrados e ir adaptando lo necesario, según las necesidades y exigencias del proyecto en cuestión.

En los proyectos en ciencia de datos no hay una búsqueda por la verdad, más bien, “es el logro de hacer que las cosas ‘funcionen’ lo que se refuerza como objetivo concreto y factible de la acción científica, y no el distante ideal de la verdad que nunca se alcanza del todo” (Knorr-Cetina, 2005, p. 59). La concretización de este imperativo, a su vez, sólo puede ser eficiente y efectiva una vez logrado un buen nivel de comunicación e interacción entre el equipo, lo que corrobora y facilita la implementación de un ambiente interdisciplinario y colaborativo.

Se nota, por lo tanto, una relación no solamente de co-dependencia entre los distintos equipos del proyecto, sino que también un cierto grado de fusión (pero no transdisciplinaria) entre los distintos conocimientos convivientes alrededor del proyecto. Es preciso que el equipo técnico (profesionales pertenecientes a la esfera datos) comprendan mínimamente las matices y especificidades compartidas por el tema-problema a dedicarse. ¿Qué significa una variable en relación a otra? ¿Como ellas se combinan en el mundo real? ¿Cómo este problema viene siendo abordado por otros trabajos científicos con o sin procesamiento de datos? Son ejemplos de preguntas esenciales para consolidar la optimización de los modelos utilizados, garantizar el buen funcionamiento del preprocesamiento, y por consecuencia, también de los resultados involucrados en la investigación.

Con el pasar del tiempo (de 12 a 24 meses), luego de las horas dedicadas al proyecto y a las diversas reuniones con especialistas, los científicos de datos terminan incorporando el vocabulario conceptual adyacente a la temática trabajada en el proyecto. Se desenvuelven en un tipo de profesión súper específico, especializados en el procesamiento de datos, y, en alguna medida, son un experto en las temáticas que involucran los datos que están trabajando.

En el inicio trabajar en este proyecto fue muy complicado para mí. Yo no entendía nada ni como pasaban las cosas al interior del Senado. Con el tiempo fui entendiendo hasta la ritualidad de las comisiones y audiencias. De tanto analizar las transcripciones de las secciones de la cámara y tratar de modelar los discursos y categorizar las temáticas de las hablas, para la implementación del programa interno que mantenemos la operación en el Senado, me torne el ingeniero que más sabe de política, leyes, recursos do Brasil [en tono de broma]. Pero, hablando en serio, nosotros, en un determinado momento, no lograbamos avanzar en el desarrollo de los algoritmos hasta que todo el equipo técnico por detrás de este proyecto tomó

un curso ofrecido a los funcionarios del Senado. Yo soy doctor en sistemas, pero tengo diversas especializaciones, en ingenierías de sistema, claro, mas tambien en temas legales, democracia y hasta en lingüística, por función de este proyecto. (ULY - 53/2023 - Brasil - mi traducción)

Los especialistas, a la vez, necesitan aprender el vocabulario de los desarrolladores, además de saber las potencialidades y límites del funcionamiento de las automatizaciones. Así, pueden ir proponiendo, por ejemplo, nuevas combinaciones de métodos que podrían ayudar a llegar a los resultados esperados. Un conocimiento limitado de las capacidades del procesamiento de los datos podría impactar directamente en la producción del conocimiento o producto deseado. En relación al análisis de los resultados de la investigación, gran parte de la responsabilidad de estos profesionales, depende, también, de su capacidad de conversar con el equipo de desarrollistas. Ser capaz de reconocer patrones y entregar valor a los índices, métricas, o parámetros encontrados a través del desarrollo de los datos, como evidencia uno de los entrevistados:

En un principio, como yo venía del mundo teórico, empecé trayendo la discusión con respecto a la lógica, matemática y modelo de análisis de argumentos. Pero después sentí la necesidad de entender la parte de método, en la creación de algoritmos para algunos análisis, como por ejemplo la estructura de lingüística con estructura lógica... Pero no, no sé cómo programar este algo. Para que haga una búsqueda automática de una estructura lingüística, ejemplo, yo conocía centenas de teoría y por mi contacto con investigaciones de esta naturaleza, en un minuto supe que, en este caso lo podía automatizar. Hasta hice un curso para entender Pipeline de la programación, me capacite para buscar Software...Pero sinceramente, para qué está gastando tiempo en algo que hay otra persona que lo hacen perfecto y que si tú logras tener una buena comunicación no tiene para que estar haciendo? Trabajar colaborativamente con el ingeniero o ingeniera, creo ser la mejor solución para el desarrollo ideal de los proyectos e investigaciones con este enfoque. (CLP - Entrevista 8/2022 - Chile)

Así, como se observa a partir de las entrevistas, conjuntamente, como en una conversación, universos prácticos y teóricos muy distintos comienzan a interactuar y apoyarse. Los vocabularios y visiones de ambos mundos pasan a integrarse de manera simbiótica y natural. Con respecto a esto, se hizo evidente que tras el contacto de los científicos de datos con los especialistas vinculados a las ciencias sociales, comienzan a interesarse por el uso ético de los datos. La preocupación por las implicaciones éticas de su trabajo (transparencia,

responsabilidad, explicabilidad de los procesos, y seguridad de los datos y sesgos de algoritmos) pasa a ser más constantes durante el ejercicio de sus funciones.

Desde el punto de vista de los especialistas, se percibe un fuerte involucramiento con la defensa de inclusión de metodologías computacionales y de automatización en sus procesos investigativos. Traen consigo impregnada la necesidad de trabajar los datos de manera computacional, lo que de alguna manera los llevan a aprender, por lo menos en un nivel básico, a hacer uso de las herramientas de procesamiento computacional de datos, por ejemplo, o bien empiezan a involucrarse más activamente de la fase de gestión de datos, transitando más libremente por las metodologías digitales, aunque sin dejar de lado las técnicas de investigación tradicionales cuando es pertinente.

Abajo, parte de la conversación con uno de los profesionales en ciencias sociales, posgraduado en ciencias políticas más solicitado y respetado con respecto a la implementación de oficinas de datos en Brasil y varias partes del mundo:

La incorporación de los procesos de la ciencia de datos a los estudios sociales no es una cuestión baladí, porque es un lugar que mezcla muchas cosas: estadística, informática, matemáticas...[...] verás, lo que hace que una persona sea un científico social...El científico social bucea en el mundo, que es lo que solemos llamar metodología. Saber bucear y explicar por qué buceas, con qué herramientas, la trazabilidad de tu buceo... Pues bien, yo me di cuenta de que con lo que llamamos ciencia de datos tenía esta capacidad de bucear en el mundo de una manera disciplinada, también, pero con las herramientas de la informática y de la estática, en general, generaba una potencia absurda para mis estudios y propósitos de investigación. Porque con la informática, pude replicar mi inmersión, sólo que desde la perspectiva y con la ayuda de la automatización. Me di cuenta de que, con un pedacito bien seleccionado del mundo, podía hablar con más propiedad del mundo entero. El mundo entero entre comillas. Entonces, ¿por qué no conectar estas cosas? El acto de hacer ciencia con el acto de automatizar cosas y con la capacidad de tomar partes del mundo, sabiendo seleccionarlas y analizarlas de una manera a menudo automática, podía mejorar mi capacidad de hablar del mundo. Así que creo que, en términos de lo que entendemos hoy en día, me he dado cuenta de que esto está muy cerca de lo que llamamos Ciencia de Datos. Esta vinculación de gran parte de la estadística con la informática y con los supuestos de la informática y la profundización de las capacidades de las ciencias sociales ¡es muy buena! En mi caso, no prefiero uno u otro; cuando funciona, utilizó una combinación de todos los métodos científicos que son compatibles y están disponibles. Por supuesto, entre aquellos con los que yo o mi equipo estamos familiarizados. (WYM - Entrevista 38/2023 - Brasil - Mi traducción).

Es justamente por la incorporación de prácticas y de un vocabulario computacional a la hora de investigar, ocurriendo al interior del ecosistema big data que encontrase un segundo hallazgo de esta investigación

#### ***8.5.4 Tolerancia y confianza en la imprevisibilidad***

La investigación social en, con y a través de la ciencia de datos crea un entorno en el que las técnicas y herramientas de procesamiento automático de datos, a menudo generan resultados efectivos en términos probabilísticos y sujetos a validación a través de numerosos parámetros y rigurosidad metodológica al seleccionar modelos de procesamiento, pero que no pueden explicarse por completo. Esta imprevisibilidad, y en muchos casos la imposibilidad de explicar la lógica interna de los resultados, se encuentra en lo que podría denominarse la zona gris del procesamiento de datos, donde son las máquinas las que operan y encuentran patrones, relaciones y probabilidades que dan lugar a los resultados.

Paralelamente, el uso de técnicas de procesamiento computacional de datos permite una inherente flexibilidad y adaptabilidad a la parte donde se producen la información de la investigación. Esta mayor tolerancia a la incertidumbre, a su vez, permite y autoriza a los científicos a explorar patrones, inferencias y resultados que no podrían haber sido detectados a través de las rutinas de investigación tradicionales de las ciencias sociales, por ejemplo. Es como si los científicos involucrados en este contexto de investigación, se permitieran un mayor grado de libertad con relación al testeo metodológico, a la construcción de inferencias y a la experimentaciones a partir de los datos.

Es chistoso, porque ahora después de 5 años trabajando juntos, estamos siempre bromeando que Fulano [su colega de investigación, Especialista en Ciencias Sociales] se volvió menos cuadrado, más creativo. Porque la gente de lingüística, sabes, son muy estructurados. Así que ahora él siente más confianza de probar, y preguntar: “Oye, mira, tengo esta data, ¿qué podríamos hacer? Está esta idea, esto sí, esto no”. Evidentemente yo, la menos volada, voy viendo la viabilidad de sus ideas. Soy la que digo: “Eso no se puede, no hay datos, no hay controles (...) no hay datos”. Eso. Como un poco la que aterrizó. (MPR - Entrevista 36/2022 - Chile)

Tener la capacidad de desarrollar mayor libertad de experimentación, no obstante, no es sinónimo de no rigurosidad metodológica. Si bien está permitido “jugar con los datos”, un proyecto serio en ciencia de datos no se esquivo de poseer una robustez lógica en cuanto a la elección de los modelos y algoritmos capaces de posibilitar las condiciones estructurales para que este juego acontezca. Del diálogo anterior, se profundiza:

\*Las frases rosadas, evidencian el papel del especialista, mientras que las verdes, del científico de datos.

“Tenemos una forma de trabajar en que es como: “Ya, *okay*, ¿qué queremos medir?” Claudio me dice: “Pucha, quiero medir la intención del hablante”. Y yo: “Ya, *okay*, pero, ¿qué teorías tenemos para eso? “Tenemos esto, Palmer, de la gramática”. Y yo le digo: “Ya, pero necesitamos entonces identificar las categorías, yo necesito que tú me digas más o menos cómo podemos agarrarla desde el texto”, y me dice “mira, tenemos como estas construcciones gramaticales y esto, ¿podemos hacer algo automático?”. Y yo le digo “sí, mira, con esta herramienta de Python podemos más o menos ver qué tipo de verbo, qué esto, esto otro”.

Entonces después yo le mando planillas de textos enormes y él va haciendo codificaciones manuales que después yo trato de programar, y se las mando de vuelta y él las corrige y vamos viendo qué cosas, qué marcadores funcionan, qué marcadores no funcionan, y vamos así en un trabajo que va de aquí a acá, y él ya más o menos entiende qué cosas se pueden programar y qué cosas no, y yo voy aprendiendo cada vez más como de lingüística y de gramática.

A través de esta conversación se puede comprobar una rutina de actividades bien definida y la presencia de pasos lógicos a seguir, como se espera de cualquier investigación. Observamos que hay una preocupación por el tipo de modelo y método que puede ser implementado a partir de la base de datos disponibles, así como con el soporte de las teorías que circunscriben los posibles hallazgos del estudio, y la presencia de un proceso de constante revisión y retroalimentación. Finalmente, la última línea nos habla, una vez más, del trabajo interdisciplinario inherente al modelo de investigación en ciencia de datos.

Con esto en mente, la tolerancia a la imprevisibilidad y al grado de indeterminación puede explicarse por varias razones. Una primera explicación es la naturaleza intrínseca al procesamiento automático de grandes datos digitales, y la manera en que se conduce una investigación en ciencia de datos; la que considera, recordemos, una diversidad de datos, el

proceso de aprendizaje automático que involucra las soluciones algorítmicas, y la interactividad entre los diferentes agentes humanos y no humanos imbricados en este tipo de investigación.

Con respecto a la diversidad, variedad y grandeza de los datos, se torna difícil prever el comportamiento de procesamiento de la máquina. Ya que las famosas corridas y calibramiento de los modelos son hechos en función de los propios datos. El científico de datos elige los modelos en base a las categorías que los especialistas consideran más relevantes. Estas son como pistas iniciales para la máquina, que a su vez tiene agencia y libertad de partir de la data de entrada fortalecida, y trazar por cuenta propia nuevos patrones de comparaciones, tendencias y correlaciones más probables al conjunto de datos fornecidos.

Es a partir de este movimiento que, de igual manera, se puede corroborar la premisa defendida a lo largo de esta investigación, en cuanto al rol fundamental de actores humanos y no humanos. En el sistema sociotécnico de investigación social en ciencias de datos ambos actores ejercen una influencia determinante en el proceso de fabricación del conocimiento y de productos, a partir de su variedad, complejidad y poder de adaptación.<sup>108</sup>

Volviendo a lo anterior, se debe entender que la posibilidad de trabajar con una suerte de gran diversidad de datos, agrega una capa adicional de complejidad e imprevisibilidad a este tipo de estudios, ya que los datos pueden tener diferentes estructuras y formatos. Ya combinarlos es un proceso de imprevisibilidad e innovación, como se explica en la siguiente entrevista a una especialista en ciencias sociales:

Para uno que no está tan familiarizado con la cosa toda, como yo en este momento, y va teniendo contacto por primera vez con las posibilidades *big data*, se queda realmente muy impresionado. Una cosa es analizar y procesar no se, un único tipo de corpus textual, como las entrevistas. Tú haces la descarga todo este material, un único tipo de archivo, por ejemplo miles y miles de PDF, para después hacer un procesamiento y análisis de lenguaje natural aplicada a los textos. Hasta ahí ok. Normal. Pero otra cosa es cuando uno de repente se ve en un proyecto como este, que más allá de analizar el texto de las entrevistas, mezcla los análisis de texto, al procesamiento de los audios de estos textos, vía entrenamiento de máquina. ¡Es sorprendente!

Estamos logrando detectar un patrón de tono de voz (si más agudo, si más graves), o la velocidad de voz (mas lento, mas bajo, mas acelerada la

---

<sup>108</sup>Es como diría Knorr-Cetina: “La cuestión no es, como nos dice la terminología de Von Poerster, la construcción del “orden” a partir del desorden (indeterminación, azar), sino el surgimiento de la organización, en tanto definida por un aumento en la complejidad o diferenciación del sistema. [...] Sólo si vemos la indeterminación como algo que se traduce en una mayor organización o complejidad, y no en orden, podemos definir este efecto como un aumento de información para el sistema y comprender el poder adaptativo que se deriva de esa organización” (2005, p.73).

cadencia) y relacionarlos a las palabras, expresiones y contexto de habla. Esto es algo muy nuevo para nosotros porque ahora el discurso puede ser analizado no solamente por el contenido, sino también por la fonética y el estudio de las emociones en el habla. Están saliendo cosas verdaderamente muy interesantes que a ninguno de nosotros [humanos] nos había dado a pensar. recién estamos abriéndonos para este tipo de estudios.[...]

Más, para ser bien honesta, aun no nos atrevemos a publicar nada más específico sobre esta investigación, mostrar resultados y como estamos haciendo, etc, porque para ser bien honesta, nosotros, a pesar de tenermos resultados que conversan con muchas de las teorías clásicas de la disciplina, no logramos entender bien bien porque los modelos que creamos están generando estas inferencias, me entiendes.

Uno de nosotros tuvimos la idea de ver con el [ingeniero de computación contratado] se había la posibilidad de unir las voces con los textos y ver qué salía, porque recién uno de nosotros había leído un paper de un grupo de EUA que hacía esto con las grabaciones y las transcripciones de los juicios de los condenados a muerte el agun estado de allá. Él, entonces, nos hizo un par de búsquedas, encontró unos modelos ya disponibles para uso, hicimos unas pruebas, y dos semanas después teníamos los “primeros resultados”. Sería bacán mostrar a nuestros pares este nuevo escenario de estudio en la lingüística, pero yo no puedo pedir que él [ingeniero de computación] me escriba estos papers, ¡me mata! yo aun no lo puedo escribir porque recién entiendo lo que estamos haciendo. Así que me contenté en publicar este texto por lo cual me encontraste, que como tú mismo dijo es más teórico, más reflexivo. (CRS- Entrevista 15/2023 - Brasil - mi traducción)

Escenarios como este crean una especie de obligación para los investigadores a estar preparados para lidiar con datos que son muy diversos, no estructurados, ruidosos y con una amplia gama de formatos. Haciendo que sus especialistas no técnicos (ingenieros, TIs y científicos de datos), tengan mayor dificultad de comprender, cabalmente, el mecanismo por el cual originan sus resultados. De todas maneras, sus capacidades de explicar el proceso de producción de información son significativamente reducidas, lo que puede generar un desconforto a los especialistas en general.

La falta de un control riguroso en la producción de información a través de la intervención humana es una novedad para las ciencias sociales. A pesar de que los investigadores, especialistas y científicos de datos eligen métodos y modelos para procesar sus bases de datos como *redes neuronales* o *árboles de decisión*, no pueden controlar ni explicar todos los procesos que transforman los datos en información. A diferencia de los métodos tradicionales de la investigación social, que se basan en la selección cuidadosa de una muestra representativa, y en la aplicación controlada de métodos conocidos para reducir la variabilidad, el uso de tecnologías big data permite una mayor flexibilidad y adaptabilidad en el análisis de los datos.

La segunda posible explicación para cierto grado de indeterminación en este tipo de contexto investigativo, es el uso de las técnicas de machine learning y la utilización de los algoritmos de aprendizaje automático. Los algoritmos de aprendizaje automático utilizados en la ciencia de datos tienen la capacidad de adaptarse y aprender de los datos en lugar de depender de suposiciones predefinidas. Esto significa que pueden encontrar patrones incluso en situaciones impredecibles y no lineales. Los modelos de aprendizaje automático pueden ajustarse a los datos en lugar de depender de supuestos rígidos. Pueden inclusive completar informaciones faltantes a una base de datos a través del entrenamiento de máquina no supervisado.<sup>109</sup>

La tercera posible explicación tiene que ver con el intercambio entre los diferentes profesionales y especialistas, provenientes del ambiente interdisciplinario de la investigación: La ciencia de datos a menudo implica un proceso iterativo en el que los investigadores exploran los datos, ajustan modelos, realizan pruebas y refinan su enfoque. Esto permite descubrimientos a medida que se avanza en la investigación y se adapta a la imprevisibilidad de los datos.

La tolerancia y la confianza en la imprevisibilidad sumada a este importante nivel de interacción entre los diferentes actores, como consecuencia, potencializa la producción de descubrimientos e innovaciones. Son cruciales para la investigación en ciencia de datos porque no se trata apenas de aceptar las reglas del juego y conformarse con la forma como se produce análisis de información en este tipo de investigación, sino que también significa dar espacio a la fabricación de innovaciones: se acoge, e incluso se esperan resultados no esperados.

La imprevisibilidad en la investigación en ciencia de datos que lleva a la producción de un ambiente investigativo innovador se manifiesta a partir de unas cuantas maneras. La siguiente tabla (28) trata de demostrar de manera sucinta estos diferentes aspectos.

---

<sup>109</sup> Ver glosario

**Tabla 28.** Dimensiones potencializadoras de innovación en proyectos de investigación en ciencia de datos

Aspecto	Descripción de cómo se causa innovación
<b>Interdisciplinaridad</b>	La imprevisibilidad en la ciencia de datos fomenta un ambiente innovador a partir de la interacción interdisciplinaria entre los diferentes profesionales involucrados, que a raíz de esta interacción, pasan a estar más abiertos a la exploración de nuevos enfoques y métodos, lo que conduce a la generación de ideas innovadoras.
<b>Diversidad de datos (condición big data)</b>	La variedad y magnitud de datos es considerada como oportunidad para la complejificación de los resultados y descubrimientos. También fomenta el aprendizaje constante, lo que promueve innovación y la mejora continua de las habilidades y conocimientos del equipo en general.
<b>Imprevisibilidad</b>	La tolerancia a la imprevisibilidad permite que los investigadores acepten resultados inesperados como una oportunidad para la innovación y el avance significativo e inesperado en la investigación.
<b>Testeo de metodología</b>	Los investigadores en ciencia de datos aceptan la inevitabilidad de los constantes ajustes metodológicos. La tolerancia al testeo de los modelos y metodologías ideales a cada proyecto, también reconoce la inevitabilidad de los errores (o no aciertos) en el proceso de investigación: son abordados mediante técnicas de limpieza y validación de datos, que muchas de las veces se constituyen como nuevos resultados, inesperados e innovadores.

En resumen, la ciencia de datos se distingue por su capacidad para lidiar con la imprevisibilidad inherente a los datos masivos y complejos. Los investigadores en este campo deben estar preparados para adaptarse a resultados inesperados y a menudo utilizan enfoques estadísticos y de aprendizaje automático para extraer conocimientos útiles de situaciones impredecibles. La imprevisibilidad se considera una característica deseable, ya que puede llevar a descubrimientos innovadores y a nuevas perspectivas en la investigación científica y social.

También reitero, a luz del pensamiento de Knorr Cetina (2005), que cierto grado de incerteza en cualquier investigación es esencial. Hay una nota de pie de página en su libro, página 72, donde la autora explica la importancia de la indeterminación en los procesos de construcción del saber científico biológico, que según mi opinión, son similares a la producción del conocimiento científico que he podido observar en los laboratorios en ciencia de datos.

Recientes desarrollos en la teoría de los sistemas autorregulados (y de la tennodinámica) sugieren la interpretación opuesta, a saber: que esa indeterminación es un prerrequisito necesario para la adaptación progresiva y organizada y, en consecuencia, para la supervivencia y el cambio reconstructivo de estos sistemas (Knorr-Cetina, 2005, p. 72).

Para cerrar el aspecto de la innovación, definiendo que el grado de indeterminación, junto a la naturaleza específica y la magnitud de los datos trabajados por la ciencia de datos, es lo que provoca el grado de libertad científica suficiente para hacer girar y crecer la rueda del descubrimiento y de la innovación científica vinculada a los métodos, las soluciones o resultados provenientes, justamente, del imperativo y el deseo por incidir, resolver problemas y experimentar este *nuevo mundo*:

Por una parte, está la capacidad de la ciencia de construir información ‘nueva’, esto es, de producir ‘innovación’. Por otro lado, la ciencia es aparentemente cada vez más capaz de construirse y reconstruirse a sí misma en respuesta a problemas que la desafían, al dar soluciones a esos problemas, lo cual, supongo, es lo que queremos decir cuando hablamos de éxito en ciencia. Ambas capacidades son aspectos de proceso de complejización, que en el sentido de Shannon corresponde a un aumento de información. Pero, como hemos visto, sin indeterminación no podría haber ese aumento de la información. Esa indeterminación parece ser nada más que los grados de libertad utilizados por el sistema para una reconstrucción de sí mismo por la absorción de problemas. La cuestión no es lamentar la existencia de la indeterminación sino verla como una parte constitutiva del aumento de conocimiento, definido por un aumento contextualmente relevante de la complejidad y la variedad. Así, una cierta indeterminación en las selecciones de laboratorio parece ser un *sine qua non* de surgimiento de información nueva (Knorr-Cetina, 2005, p.75).

No podría, de igual forma, dejar de hablar de la tolerancia al error, e incluso su importancia. En la investigación en ciencia de datos, existe una mayor tolerancia a los errores debido a la propia naturaleza de los datos masivos y complejos. Dado que los conjuntos de datos pueden contener una amplia variedad de información, algunos errores son inevitables. Son parte del proceso de la construcción de una base de datos más robusta, cuanto también del proceso de ajuste de los métodos y modelos a ser elegidos. Los investigadores en ciencia de datos, más que conformados, están preparados para lidiar con estos errores mediante técnicas de limpieza de datos y métodos de validación, haciendo que la tolerancia a los errores sea

esencial para permitir la exploración y el descubrimiento en un ambiente investigativo, y como vimos, generar igualmente innovación.

Los investigadores, estando dispuestos a explorar nuevos enfoques y métodos para comprender y analizar datos, también movilizan mayor disposición en aceptar los errores y las imprevisibilidades que conllevan. Así, se promueve la creatividad y la innovación en la resolución de problemas, en un ambiente de aprendizaje constante, lo que aumenta la confianza, estructura y mejoran los datos a ser trabajados, lo que, en teoría, también ayuda en la calidad de la información y las soluciones producidas. Muy parecido a lo que manifiesta la autora:

En los procesos estocásticos de nivel molecular, en los cuales cuanto menor el número de partículas interactuantes mayor el rol de la fluctuación, se ha mostrado que la ausencia de "errores" o indeterminación corresponde no sólo a una ausencia de innovación y por lo tanto de incremento de la información, sino a una real pérdida de información. Sin las fluctuaciones de chance, los sistemas no pueden mantenerse en un estado estacionario. Esto significa que sin la intervención de "error", azar o indeterminación en la evolución biológica, por ejemplo, todas las especies desaparecerían sin ser reemplazadas por otras. Cf. Atlan (1979: 54 ss.). Para una propagación de la idea como principio de orden relevante a la ciencia véase particularmente Latour y Woolgar (1979). Para el análisis filosófico más iluminativo que trata el tema, véase Serres (t 980). (Knorr-Cetina, 2005, p. 72)

Estos puntos resaltan la importancia de considerar y gestionar adecuadamente el error en la investigación en ciencia de datos aplicada a las ciencias sociales, y cómo esto puede mejorar la calidad de la investigación y la toma de decisiones. Abajo, presento la tabla 29 con las descripciones de cada punto positivo del reconocimiento del error en investigaciones de ciencia de datos en el contexto de las ciencias sociales:

**Tabla 29.** *El papel del error en proyectos en ciencias de datos en el marco de las ciencias sociales*

<b>El papel del error en proyectos en ciencias de datos en el marco de las ciencias sociales</b>	<b>Descripción</b>
<b>Conciencia de las Limitaciones</b>	Reconocer la importancia del error fomenta una mayor conciencia de las limitaciones inherentes a los datos y los modelos utilizados en la investigación en ciencia de datos.
<b>Transparencia en la Comunicación de Resultados</b>	Lleva a una mayor transparencia en la comunicación de resultados. Los investigadores proporcionan información sobre la precisión de los modelos y análisis, lo que permite a otros comprender las incertidumbres asociadas.
<b>Calidad de los Datos</b>	Impulsa a los equipos a mejorar la calidad de los datos. Esto puede incluir procesos más rigurosos de limpieza de datos, validación y estandarización para minimizar los errores de medición.
<b>Decisiones Basadas en Datos</b>	Conduce a la toma de decisiones más informadas. Los responsables de la toma de decisiones pueden comprender mejor las limitaciones de los datos y los modelos y tener en cuenta la incertidumbre al tomar decisiones.
<b>Modelos más Robustos</b>	Los científicos de datos pueden enfocarse en desarrollar modelos más robustos que sean menos sensibles a las fluctuaciones y al ruido en los datos. Esto mejora la capacidad de los modelos para generalizar y predecir con precisión.
<b>Mitigación de Sesgo y Discriminación</b>	El reconocimiento del error puede ayudar a identificar y abordar sesgos y discriminación en los datos y los modelos. Al comprender las fuentes de error, es posible tomar medidas para reducir el impacto de los sesgos y garantizar una toma de decisiones más justa y equitativa.
<b>Énfasis en la Ética de Datos</b>	La importancia del error está estrechamente relacionada con las cuestiones éticas en la ciencia de datos. Esto fomenta una mayor atención a las implicaciones éticas de la recopilación y el uso de datos, lo que es esencial en las ciencias sociales.
<b>Desarrollo Metodológico</b>	Impulsa el desarrollo de metodologías y técnicas más avanzadas para medir, evaluar y controlar el error. Esto contribuye al avance de la investigación en ciencia de datos y las ciencias sociales.

En resumen, la naturaleza impredecible de los datos desafía a los investigadores a aprender constantemente y adaptarse a situaciones cambiantes. Esto crea un ambiente de aprendizaje continuo y promueve la mejora constante de habilidades y conocimientos. Los errores se

consideran oportunidades para aprender y mejorar, lo que a su vez contribuye a un ambiente investigativo más innovador.

### ***8.5.5 La representatividad de la muestra pierde relevancia.***

La utilización del enfoque en ciencias de datos en investigaciones científicas sociales aplicadas ha llevado a un cambio en la forma en que se considera la representatividad de la población, produciendo un cambio con relación al valor de la noción de representatividad y de muestra. La investigación clásica, basada en el análisis de muestras representativas rigurosamente seleccionadas a través de técnicas estadísticas, hace que la elección de una muestra representativa (y la aplicación de métodos estadísticos) fueran consideradas fundamentales para garantizar la fiabilidad de los resultados y la validez de las investigaciones. Sin embargo, la investigación a partir de tecnologías big data tienden a intentar capturar el comportamiento de una población en su totalidad, haciendo que la noción de representatividad de la población haya perdido relevancia, cuya validez incluso se ha visto cuestionada, bajo la acusación que está quedando cada vez más obsoleta.

Las investigaciones basadas en la ciencia de datos, a través de la utilización de grandes y variados conjuntos de datos, a menudo desestructurados, trabajan a partir de la totalidad de los datos disponibles. Esto implica, en general, que no se están extrayendo subconjuntos representativos del universo de los datos, sino que se analiza el conjunto completo de los datos disponibles, cambiando así la forma en que se aborda la inferencia y la producción de los resultados relacionados a cada investigación. Tal panorama acaba por afectar o transformar la manera en la que tradicionalmente se construye el conocimiento social alrededor de dichos proyectos. Como podemos fácilmente identificar a partir de la entrevista realizada con uno de los científicos sociales especializados en datos, autor de una publicación académica perteneciente a librería Scopus, analizada:

Es interesante que preguntes esto porque esta polaridad también ha surgido con mucha fuerza en los estudios de opinión pública. Es el tema de mi investigación de máster, que acabaste leyendo en forma de artículo. Pues me pareció muy interesante y me di cuenta de que el big data y estas tecnologías de datos están permitiendo una visión de la opinión pública más inspirada en lo que aquellos autores teóricos de siglos pasados entendían por opinión pública

Bebí mucho de la fuente de Susan Herbst (1993) para apoyar mis puntos de vista críticos sobre el desarrollo de la opinión pública a lo largo del siglo XX. En sus libros, ella demuestra algunas posibilidades en la introducción de

nuevas metodologías digitales que surgen de big data. Así, por ejemplo, desde un punto de vista práctico, para ilustrar: Una encuesta de opinión pública tradicional, recogerá datos de forma muchas veces espaciada, a veces en encuestas documentales, siempre en cuestionarios, o en casos más raros, entrevistas. La opinión pública en general tiene un sesgo hacia la producción de información que se parece más a preguntas cerradas, como decimos. Y, por tanto, esta información siempre se analiza en forma de tabla o se visualiza en un formato gráfico más tradicional. En otras palabras, una visión muy estricta y extendida desde los años 30, ¡y no ha cambiado! Sigue siendo lo mismo, los mismos resultados, las mismas metodologías. Quizá se hayan perfeccionado algunos procesos estadísticos, pero en general seguimos utilizando la misma metodología de siempre.

El problema es que con este tipo de visión no podemos distinguir cómo se ha organizado la sociedad o cómo se ha formado la opinión pública hasta llegar a ese punto. Es una visión muy simplista, muy corta de miras, no en el sentido de equivocada, sino en el de borrosa.

Cuando llegamos a los estudios de opinión pública realizados a través de Internet, y podemos tomar como ejemplo los estudios de redes sociales, podemos distinguir qué grupos participan en un determinado debate, cómo están organizados esos grupos, qué grupos tienen más voz o más relevancia, quiénes son los líderes de esos grupos, etc. Entonces podemos realizar un estudio que ponga de relieve las relaciones de poder dentro de cada uno de los grupos, por ejemplo.

En la investigación tradicional suponemos que todo el mundo, la opinión de todo el mundo tiene el mismo peso, tanto Bloomberg como Bourdieu también discuten este punto. Y esta situación en la práctica, la opinión pública en el mundo real, no funciona así: Ciertas personas tienen opiniones más determinantes o más influyentes debido a la posición social o financiera en la que están insertas. Así que los sondeos de opinión pública realizados a través de las redes sociales, por ejemplo, nos permiten ver este tipo de situaciones e influencias.

Los sondeos de opinión pública tradicionales tenían este aspecto, digamos, separador, en el sentido de que las opiniones eran individuales y no se comunicaban entre sí. Cada uno tenía su propia opinión y correspondía a la opinión individual de cada persona. A través de las redes sociales, por ejemplo, podemos tener una visión relacional, y eso es muy importante. Así que ya ves, en realidad, la opinión individual de una persona sólo tiene lugar dentro de una red de opiniones, es decir, sólo tiene lugar como una relación de opiniones que forman ahí un agregado, que llamamos opinión pública. Lo cual tiene todo el sentido, porque la opinión pública es necesariamente la opinión del público, no es la opinión individual.

Así que es en este tipo de dirección que esta conclusión a la que estamos llegando, con las encuestas digitales de opinión pública, que estamos empezando a avanzar hacia la posibilidad de dejar de lado, o dejar más de lado, una visión estrictamente cuantificada hacia una visión más relacional o conversacional, porque nace de los conflictos de intereses, nace entre conversaciones, entre la comunicación entre las personas. Incluso desde el

punto de vista de la comunicación, que es mi área de origen. (PCL - Entrevista 09/2023 - Brasil - mi traducción).

Como podemos ver, el entrevistado señala que el big data y las tecnologías relacionadas con los datos computacionales digitales están proporcionando una nueva perspectiva de la opinión pública, la cual que implica pasar desde una visión cuantitativa y simplista a una más relacional y conversacional basada en las interacciones y las conversaciones entre las personas. Esto proporcionaría a la disciplina un refinamiento de sus prácticas investigativas, y como consecuencia, la producción de resultados, conclusiones y reflexiones serían más acorde con las especificidades actuales de la sociedad contemporánea.

Se demuestra acá también que una de las implicaciones con respecto a los procesos de producción de las generalizaciones hechas por las investigaciones versus las descripciones es que mientras que la investigación tradicional busca generalizar los resultados a una población más amplia, basándose en muestras representativas, el enfoque de la ciencia de datos tiende a centrarse en la descripción y el entendimiento de patrones dentro y a partir del universo de los datos. Esto significa que los hallazgos pueden ser altamente específicos a un determinado contexto de investigación en lugar de generalizables a una población más amplia.

Con respecto a lo último, a partir de las entrevistas realizadas a los científicos sociales autores de algunos de los artículos analizados, surge un debate bastante interesante con relación al no cuadramiento de las técnicas estadísticas de producción de información, si se compara a abordajes metodológicos sobretodo computacionales, que han emergido en la actualidad. Así como el entrevistado citado anteriormente, a continuación otro extracto demuestra un cierto desajuste entre las actividades que envuelven la producción del conocimiento a través de técnicas estáticas y muestrales en detrimento a las especificidades de la sociedad contemporánea. Es una observación realizada por parte de un autor de uno de los artículos de Scielo en ciencias sociales, académico y profesional en un instituto que realiza proyectos investigación social ligada al comercio y economía en Brasil:

**Entrevistado:** Aquí en la empresa tenemos un debate directo sobre la forma en que obtenemos los datos de nuestras investigaciones. Aquí también trabajan muchos ex académicos.

**Entrevistador:** ¿De qué hablas concretamente?

**Entrevistado:** De la antigüedad de nuestros métodos. En otras palabras, el modelo estadístico ya no es apropiado para nuestros tiempos. sólo por poner un ejemplo, creo que cualquier instituto de investigación debe estar pasando en estos momentos: ¡es muy difícil obtener respuestas! Que la gente coja el teléfono o responda a un cuestionario en persona en su casa. Eso ya no se puede hacer. Nadie está en casa, nadie quiere descargarse una aplicación en el móvil para responder a una encuesta que no le interesa lo más mínimo... En fin. Ah, y es muy habitual que las mismas personas respondan a varias encuestas.

También hay otra cuestión, que tiene más que ver con la estructuración del plan de muestreo. Un aspecto sociológico, digámoslo así, que es el hecho de que las encuestas, por su modelo de obtención de información, acaban subestimando ciertos estratos de la población: si utilizas el teléfono, obtienes más respuestas de personas de las clases A, B. Si no lo utilizas, si entrevistas a la gente en casa, consigues rellenar mejor las otras clases. Y esto sin duda tiene que ver con el hecho de que quizá no estén bien representados dentro de los sectores oficiales del censo. Aquí en el instituto, cuando una empresa nos llama y nos pide una encuesta fiable, ¿qué hacemos? ¡Usamos la PNAD de 2010 como plan de muestreo! Esto está desfasado para hoy. En los años 90, todavía se consideraba que una base muestral de 10 o 15 años estaba bien para ser representativa de la población. Y entonces, para cumplir con la muestra, tienes que estimar más o menos cuánto representa. Y para ello, las bases de las que partimos como universo no son tan buenas".

**Entrevistador:** Bueno, entiendo todo eso, pero ¿no se podría tomar esta gran hoja de cálculo de datos, esta tabla gigante de datos disponibles, que es el universo y hacer estadísticas sobre ese universo?

**Entrevistado:** Sí. Esto es válido y se puede hacer, pero el problema es que, para nosotros aquí, no será representativo. No significa que esas empresas que están ahí, de miles de páginas, representen lo que es la economía, ¿sabes? Aquí partimos de un registro oficial de empresas brasileñas que tiene el IBGE, que la Receita Federal pasa al IBGE. A partir de ahí tenemos el universo de empresas brasileñas, que son miles, y a partir de ahí sacas la muestra en función de la distribución de las características. Así que puedo entrevistar a las empresas y en el caso de este año había 4.000 empresas, que a su vez representan alrededor de 500.000 empresas. Así que esa es la lógica: lo sé, doy ponderaciones, puedo vincular esas características a la empresa. Esa es la importancia de la muestra.

El problema del big data es la falta de capacidad para no sesgar el estudio. Si me das una hoja de cálculo con las páginas web de miles y miles de empresas en Brasil, está bien. Pero como ya he dicho, los datos están muy sesgados. El mero hecho de que una empresa tenga un sitio web es, por ejemplo, muy restrictivo. Porque sabemos que sólo el 58% de las empresas de Brasil tienen página web. Así que de estas páginas, sólo tomó un tipo específico de página. Sí, es una estadística interesante, pero no es representativa. Es válida, pero no representativa. Así que, estrictamente hablando, cada vez que alguien se refiera a esta encuesta, debería hablar de ella, porque si no la gente piensa que estás hablando de todo, y en realidad no es así, y entonces creas otro problema, ¡que es la falta de verdad! (LLS - Entrevista 24/2022 - Brasil - mi traducción) .

Se ve que todavía existe esta gran controversia entre métodos e implicaciones metodológicas tradicionales en estadística y con el uso de ciencias de datos. Aunque el científico reconoce que el proceso de recolección de información a través de la estadística posee numerosas fallas que invisibilizan en muchas parte del proceso de la pesquisa, también está por otro lado la confiabilidad en este método estadístico como validación de los resultados como una especie de ciencia superior. Las técnicas en ciencias son válidas, pero no son consideradas representativas, lo que parece entonces quitar un poco de peso a su rigurosidad científica.

Independiente del resultado final de esta controversia, que está lejos de ser resuelta, es casi unanimidad entre los entrevistados que en la medida que la tendencia utiliza métodos computacionales para el procesamiento de información digital en la investigación social resulta, muchas veces, en una mayor precisión en la descripción de patrones y relaciones, por otro lado también pueden incitar al error y la producción de resultados sesgados, debido no respeto a la proporcionalidad de determinadas importantes variables (Anderson, 2008; Boyd y Crawford, 2012).

Como contra respuesta, los académicos optimistas defienden que justamente en consecuencia de esta característica es que el análisis de datos no estructurados puede proporcionar información valiosa, no capturada por el modelo de investigación que soporta las encuestas tradicionales.

Uno de los beneficios de la posibilidad de analizar los datos de manera no estructurada, no generalizada y no muestral, está en la promesa de alcanzar una mayor eficiencia en la incidencia y resolución de problemas debido al enfoque contextual y más específico proporcionado por la focalización en la incidencia y resolución de problemas. Aunque la representatividad de la muestra puede perder relevancia en ciertas investigaciones de ciencia de datos, el contexto sigue siendo preponderante. La interpretación de los resultados debe basarse en un profundo entendimiento del contexto en el que se recopilaron los datos y en las teorías que respaldan el proceso de validación de las soluciones, llevadas a cabo por medio de las experticias y conocimientos conceptuales proveniente de los especialistas.

Si la concentración de esfuerzos investigativos está en entender el contexto específico, las matices peculiares de la base de datos disponible y su relación con la observación de una realidad conocida, estudiada y validada por los especialistas, la rigurosidad probabilística ya

no es considerada tan esencial, ya que el objetivo final no es comprender el comportamiento de la totalidad de una población y ser fidedigna a los subconjuntos que las componen. Una investigación circunscrita en un ambiente en ciencia de datos, lo que desea, es mejorar su capacidad de manipular los datos disponibles, para conocer las maneras más apropiadas de resolver la problemática de un grupo, o un conjunto de datos específico, contextual y altamente cambiante.

El foco en el conjunto de datos, acaba provocando, por lo tanto, la necesidad de un enriquecimiento de los datos. Es decir, la necesidad de cuidar con atención los procesos que aseguren la calidad de su data de entrada, la complementariedad de estos datos y el detallamiento de los fenómenos estudiados a partir de ellos. Esta atención especial al contexto, a la especificidad de la base de datos a ser trabajada y al amplio acceso a conjuntos de datos más grandes y diversos, podría conducir a una comprensión más profunda de los patrones y las relaciones en la sociedad, lo que resultaría en un consecuente enriquecimiento de la investigación social de los fenómenos estudiados.

Lo que tenemos que esforzarnos por entender al investigar el paradigma de la investigación en ciencia de datos es que ahora, con la investigación cada vez más digital, tenemos que entender muy bien cuáles son las premisas de cada estudio. Porque la premisa de un estudio, la premisa metodológica, no siempre tiene que ser hacer una predicción o entender un problema a partir de una determinación demográficamente representativa de la población. A veces queremos analizar cómo se comportan los participantes, por ejemplo, de las redes sociales en relación con un determinado tema, y punto. En ningún momento pretendemos, queremos, necesitamos, por ejemplo, extrapolar esos datos para hacer predicciones estadísticas basadas en las muestras que tenemos, el 100% de las veces. (PCL - Entrevista 09/2023 - Brasil - mi traducción).

Un tercer beneficio apuntado que surge frecuentemente en las entrevistas es el impacto con respecto a los cambios en la toma de decisiones. La capacidad de analizar grandes conjuntos de datos, muchas veces en tiempo real, la interacción entre diferentes tipos de información y el nivel de detalle que se puede obtener a partir del procesamiento computacional de los datos, ha llevado a cambios significativos en la toma de decisiones. A esto se suma, una vez más, el imperativo de la incidencia y la resolución de problemas. Investigadores, agentes de gobierno, servicios ofrecidos a poblaciones e incluso empresas y otras organizaciones, están recalibrando la manera en cómo toman decisiones basada en datos, logrando así un impacto directo en políticas y estrategias sociales contemporáneas. Ya no se busca reflejar o entender

estadísticamente la realidad en sí misma, sino de proveer información que permita transformar esa realidad con un cierto grado de efectividad y un margen de error.

Esa predisposición a la incidencia y a la resolución de problemas, ya discutida en los apartados anteriores, más allá de ser apuntada por muchos científicos entrevistados como beneficiosa, se relativiza, a su vez, a a partir de la discusión con respecto a los riesgos asociados a estos cambios y a los puntos de atención que necesitan ser discutidos a consecuencia de estos cambios.

El primero de ellos se relaciona con la capacidad de aprovechar el poder de los datos masivos para comprender y abordar problemas sociales, pues es un desarrollo importante en el campo de la investigación social. Esto, siempre desde la óptica de la transparencia sobre los datos que se utilizan y cómo esos datos se relacionan con la población de interés, o a la parcela de la población en que deseas incidir.

Yo les digo a los presentes que cuando se analiza una base de datos por la que pasan 7 millones de personas registradas. Es una forma de hacer estadística por principio, no necesariamente por muestreo. Les propongo este ejercicio: Imaginen que hay 7 millones de personas sentadas en esta sala. Dependiendo de cómo lo analices, podrías estar cometiendo un genocidio estadístico, por ejemplo. "Voy a eliminar esta categoría indígena, porque no hay mucha gente". Bueno, acabas de matar a 800.000 personas en Brasil. He visto a gente hacer eso... "Estás matando a 800.000 personas" (WMP - Entrevista 27/2023 - Brasil - mi traducción)

Como la cita de arriba muestra, existe una preocupación en reproducir una investigación que sea seria y que contemple las especificidades muestrales de un conjunto de datos cada vez que sea necesario. Esta elección es hecha principalmente por parte de los especialistas en ciencias sociales a partir de la expertises que manejan y las unicidades vinculadas al problema-tema de las investigaciones a ellos asociadas.

Aunque no se busque una muestra representativa en el sentido tradicional, es esencial ser consciente de los posibles sesgos inherentes en los datos utilizados. Estos sesgos pueden provenir de la forma en que se recopilan, almacenan o registran los datos, y pueden afectar la validez de los resultados. Es importante que los investigadores sean transparentes sobre estos sesgos y sus posibles limitaciones.

Otro importante aspecto y foco de atención es el proceso de validación y evaluación de modelos. En lugar de depender de la muestra representativa, la validación y la evaluación de modelos son prácticas esenciales en la ciencia de datos. Los investigadores deben utilizar técnicas como la validación cruzada y la medición de la precisión del modelo para asegurarse de que sus modelos sean robustos y útiles en la toma de decisiones. No es que las investigaciones en ciencias de datos no tengan rigurosidad de resultados o no se ocupen de la solidez de sus estudios, pero los medios para llegar a ello son otros. Alias, la ciencia de datos a menudo involucra modelos estadísticos y de aprendizaje automático altamente complejos. La interpretación de estos modelos puede ser difícil y desafiante, lo que dificulta la explicación de los resultados a audiencias no técnicas, incrementando aún más el grado de rigurosidad de dichos estudios.

En consecuencia a este panorama, hay también una fuerte preocupación en cuanto al acceso y a la producción de datos de alta calidad, el cual es un desafío importante. En muchos casos, los datos disponibles pueden estar incompletos o contaminados, lo que dificulta la obtención de resultados precisos y confiables.

Para terminar y resumir, la siguiente tabla sintetiza las potencialidades y los desafíos relacionados a la pérdida de relevancia de la representatividad de la muestra en la producción de conocimiento social a través de investigaciones en ciencias de datos.

**Tabla 30.** *Potencialidades y desafíos relacionados a la pérdida de relevancia de la representatividad de la muestra*

Aspectos	Potencialidades	Desafíos
<b>Datos</b>	Ofrece una imagen más completa y detallada.	Privacidad y protección de datos.
<b>Descubrimiento de patrones inesperados</b>	Identificación de nuevos problemas.	Sesgos en los datos.
<b>Enfoque en problemas prácticos</b>	Soluciones más efectivas y eficientes.	Acceso a datos de alta calidad.
<b>Cambios en la toma de decisiones</b>	Toma de decisiones más informadas.	Falta de teoría y contexto.
<b>Ética y privacidad de datos</b>	Garantiza el respeto a la privacidad.	Interpretación de modelos complejos.

<b>Sesgos en los datos</b>	Ofrece información incompleta o contaminada.	Capacitación y habilidades.
<b>Acceso a datos de alta calidad</b>	Mayor confiabilidad en los resultados.	Costos y recursos.
<b>Falta de teoría y contexto</b>	Enfoque en la objetividad y la imparcialidad.	Interdisciplinariedad.
<b>Interpretación de modelos complejos</b>	Mayor precisión en la predicción.	Seguridad de datos.
<b>Capacitación y habilidades</b>	Capacidad para trabajar con datos masivos.	Cambio en las dinámicas de trabajo.

Podemos percibir que en las investigaciones de ciencia de datos, la noción tradicional de representatividad de la muestra puede ser reemplazada por un enfoque en la descripción de patrones dentro de los datos y la validación de modelos. Sin embargo, esto no significa que la representatividad sea irrelevante, sino que se debe entender y abordar de manera diferente en función de los objetivos y las características de los datos en cada estudio.

En general, el cambio hacia la ciencia de datos en los estudios sociales ofrece muchas oportunidades, pero también presenta desafíos. Los investigadores deben adaptarse a nuevas metodologías, ser conscientes de cuestiones éticas y de privacidad, y estar dispuestos a colaborar en equipos interdisciplinarios. La capacidad de aprovechar el poder de los datos masivos para comprender y abordar problemas sociales es un desarrollo importante en el campo de la investigación social, que, que al fin y a cabo, necesita ser conducida idealmente a partir de la integración de estos dos mecanismos de obtención de información, cosa que hablaremos con más profundidad adelante en el capítulo 10: "En la versión extranjera de mi artículo, hago hincapié en que los modelos no deben verse como mutuamente excluyentes. Deberían verse como complementarios porque lo que uno tiene que probar, el otro puede aportar nuevas posibilidades y éxitos ". (PCL - Entrevista 09/2023 - Brasil - mi traducción).

### **8.5.6 Cuantificación y visualización**

La última implicación metodológica a la investigación social a través de y con tecnologías en ciencia de datos, es el mayor uso de métodos de investigación cuantitativa y la atención especial a las técnicas de visualización de datos.

Con respecto a la primera, es posible observar que la aplicación de métodos de procesamiento computacional avanzados de datos, la gestión de grandes bases de datos y el dominio de lenguajes y técnicas de programación son esenciales para extraer información significativa de los estudios que se pretende llevar a cabo. Esto ha llevado a dar una mayor importancia a este tipo de habilidades técnicas, que en general son escasas y más aún en las ciencias sociales. Como resultado, varios investigadores se han visto en la necesidad de capacitarse para adquirir estas nuevas habilidades técnicas.

En una conversación informal con uno de los científicos de datos de GobLab, comentábamos sobre lo difícil que había sido para el laboratorio lograr encontrar y contratar a alguien con la formación que tenía él para conducir la parte de procesamiento y desarrollo de las soluciones involucradas en los proyectos activos de dicho laboratorio. Le comenté que decíamos que si uno de nosotros (científicos sociales) tuviera una formación “más metodológica” y supiera programar, podríamos fácilmente realizar este trabajo, ganar más por esto y aun solucionar unos de los principales desafíos del laboratorio. Lo que me contestó él:

Ahhh, muy bien, pero el primer problema es que, si hay una persona que está desde el área de sociales o desde cualquier área diferente, no tan dura del modelo, uno dirá “listo, ya generó la curiosidad sobre tal tema, yo tengo las entradas y yo tengo unas salidas, y yo tengo un modelo”. Entonces, ellos dicen “oye, yo quiero meterme, no estar como tan de la parte de la entrada ni tan de la parte de la salida validando, sino que yo quiero estar acá, en el medio, procesando”. Entonces, yo le diría “ah, bueno, tranquilo...Pues primero vas a tener que aprender matemática, específicamente álgebra lineal, optimización, estadística, primero que todo, dura y pura, ¡y luego un poquito de programación!” Cuando sepa álgebra lineal, optimización, estadística y programación, esta persona de ciencias sociales tiene trabajo garantizado de por vida y mucho billete. (RNN - Conversación Informal - 2022 - Chile)

Se percibe, pues, que para insertarse en un sistema de investigación en procesamiento automático y computacional de los datos, es preciso poseer determinados conocimientos, o tipos de saber científicos que por lo general no están asociados a las carreras y a la formación en ciencias sociales, vuelve aún más específico y selecto el profesional que logra, por lo menos, dialogar con los tecnicismos y especificidades de este tipo de saber.

A pesar del fuerte énfasis, especialmente por parte de los especialistas en ciencias sociales y otras disciplinas, en abordajes que integren procesos de recopilación y análisis de información mixtos conocidos como la integración de metodologías cuantitativas y cualitativas, es innegable que en este tipo de investigación el proceso de producción de información cuantitativa ocupa un lugar preponderante. Este proceso cuantitativo destaca por sí mismo.

El proceso de recolección de información cualitativa, que incluye entrevistas, grupos focales, seguimientos, observación participante y, en ocasiones, incluso encuestas (aunque tradicionalmente se consideren cuantitativas en las ciencias sociales), complementa la información y los resultados generados a partir de enfoques de estudio predominantemente cuantitativos. Estas metodologías cualitativas se centran principalmente en las etapas iniciales del proyecto, como el mapeo de actores clave en la investigación y la comprensión de las opiniones de futuros usuarios del sistema a desarrollar.

También desempeñan un papel importante en la fase final de la investigación, facilitando la introducción de soluciones y mejoras en la usabilidad de los productos desarrollados en la rutina diaria de los grupos destinatarios. En la secuencia presenté una nota de mi diario de campo, luego después de una reunión con el equipo gestor del proyecto fondef SAMSARA, luego después de mi contratación y integración a este nuevo proyecto:

Hoy tuvimos la primera reunión con el equipo Cualitativo del Proyecto SAMSARA. Todo bien en la reunión, las personas somos las mismas. Lo que me llama la atención, y es hasta chistoso para mi, es que la actividad principal destinada a este equipo es la ejecución de un survey con los trabajadores/usuarios de imágenes satelitales en el estado chileno. ¡Un survey! Yo soy la coordinadora de la investigación cu.a.li.ta.ti.va y estoy encargada de llevar a cabo una clásica metodología cuantitativa para las ciencias sociales ¡Ironías de un campo muy exacto y duro! (Notas de campo, 14 de enero de 2022 - GobLab)

Por más que los científicos sociales definitivamente no sean los contratados para “rodar” los datos de los proyectos de investigación del que son parte, es cada vez más nítida la sensación de que una mínima capacitación en lenguajes de programación, estadística y matemática es un requisito implícito para seguir ganando espacio en este medio. Se necesita incluso para facilitar la comunicación entre las distintas partes del proyecto, por ejemplo para dialogar con la parte metodológica de los proyectos, poder cuestionar, sugerir, aportar y comprender mínimamente las decisiones tomadas por los equipos técnicos. Por ello, es un conocimiento cada vez más esencial, aunque no lo utilices en la práctica, pues garantiza la transitividad entre las diferentes estaciones de la investigación.

Históricamente, muy pocos cursos de ciencias sociales han enseñado un análisis estadístico real y en profundidad, cómo hacer buenos cuestionarios, analizar, etcétera. Pero creo que hoy en día no hay excusa. Nadie sale de la universidad sabiendo todo lo que necesitará para trabajar en la vida real. Por desgracia o por suerte, tenemos que ir un paso más allá y aprender las nuevas metodologías, que hoy en día son las metodologías digitales. La gente de ciencias sociales tiene que empezar a estudiar R, a estudiar Python para hacer mejores análisis. No hay manera de evitarlo. O sea, para mí, lo digo con un poco más de sitio para hablar, que, hombre, pensar en analizar datos sin Python, a mí, me parece una cosa ridícula, en serio. porque si no lo usas, hombre, no estás siendo inteligente. Tanto para recoger como para analizar los datos. Es hora de que este científico social aprenda, porque si no estaremos en una situación de rehenes. Si nos quitamos ese trauma de los números que tenemos, somos perfectamente capaces de aprender unas metodologías digitales, unos pasos de programación y ¡listo! (PCL - Entrevista 09/2023 - Brasil. Mi traducción)

Algo parecido ocurre con las técnicas de visualización de datos, que también se ha convertido en una herramienta fundamental tanto del proceso científico, como de la toma de decisiones y la comunicación de los resultados. Proporcionar visualizaciones intuitivas y efectivas de los resultados de una investigación es esencial para comprender y analizar los complejos patrones y tendencias que se identifican en estas grandes bases de datos. Estas visualizaciones facilitan la comunicación de resultados a audiencias más amplias y diversas, lo que es esencial para la toma de decisiones informadas en contextos empresariales, gubernamentales y sociales. Además, la visualización de datos puede ayudar a identificar errores y deficiencias en los datos, lo que a su vez permite mejorar la calidad de la investigación y la precisión de los resultados.

Sin embargo, es importante tener en cuenta que la visualización de datos también tiene sus limitaciones. La elección de la forma de visualización puede influir en la interpretación de los datos, y los gráficos pueden ser utilizados de manera engañosa para transmitir una imagen incorrecta de los resultados. Por lo tanto, es importante utilizar técnicas de visualización variadas para proporcionar una imagen más completa y precisa de los resultados.

Como estas investigaciones están orientadas para la incidencia en el mundo, para la mejoría de una función, optimización de algún proceso de análisis y de toma de decisión y control sobre el ambiente social, la etapa de “venta/entrega del producto” cobra importancia preponderante. Es importante saber comunicar la relación de las soluciones desarrolladas, al usuario final, a los apoyadores y a los agentes de financiamiento.

Es clave comunicar la relevancia del proyecto para la ciudadanía o el servicio público, si estamos hablando, por ejemplo de iniciativas gubernamentales, como suele ser en los países y proyectos de América Latina. Otra razón por la que la visualización de los hallazgos y soluciones aplicadas es de especial importancia. El procesamiento computacional de los datos trae consigo una especificidad relativamente grande para los no letrados en programación, desarrollo de sistemas y lenguajes computacionales. Para que la efectividad de la aplicación de proyecto sea garantizada, es necesario que todas las partes involucradas en el proyecto comprendan mínimamente el desarrollo y ejecución interna de la investigación.

Es importante no solamente transparentar los procesos de investigación, los modelos elegidos, los porqué de la mayor parte de las decisiones tomadas, pero también que el usuario final de las soluciones proyectadas pueda entender los resultados, visualizar la potencialidades de los proyectos y comprender la importancia de su aplicación y lo que cambia partir de la introducción de estos nuevos conocimientos y prácticas derivadas.

Después de más o menos 10 meses de iniciada esta investigación, fuimos a presentar los resultados al gabinete. Estábamos extremadamente contentos, hasta un poco ansiosos, porque habíamos hecho un trabajo excepcional, y los resultados obtenidos fueron mucho más significativos que los resultados esperados, en fin. Lo que pasa es que después de la presentación, hay un silencio total. Nosotros del proyecto nos mirábamos y preguntábamos internamente que había pasado, ¿porque nadie nos felicitaba! Hasta que alguien tuvo el coraje de levantar la mano y decir que entendía el canva del proyecto, pero que no había entendido el propósito final de aquello. Tuvimos que replicar los hallazgos del proyecto, y hacer malabarismos con nuestros números para que la audiencia pudiera realmente entender lo que pasaba, y en

este proceso perdimos la capacidad de sensibilizarlos. La respuesta por parte de sus gestores no fue la que esperábamos, la que tenía que ser...

Pero hay que salir adelante, ¿cierto? Ese día nos trajo un aprendizaje, una lección como equipo: necesitamos aprender a traducir nuestros conocimientos para un público "ignorante". Llegamos a un acuerdo interno entre el equipo. Dos de nosotros hicimos un curso en línea de *power bi* [programa que posibilita el diseño de paneles de información a partir de datos] y ahora tenemos a alguien del departamento de estadística trabajando directamente con nosotros en la fabricación de los materiales de comunicación y visualización de los datos generados. Somos cada vez más eficientes en vender nuestras cosas!. (Conversación grupal equipo Tribunal Superior Electoral - 2023 - Brasil - mi traducción)

Corroboramos la importancia de la forma como se comunican, proyectan y evidencian los resultados de las investigaciones a la sociedad, o al pequeño grupo de agentes a quien está direccionado. Incluso porque tales proyectos, como ya mencionamos anteriormente, tienen la vocación de incidir en la realidad, sea por medio de optimizaciones de procesos de trabajos, sea a través de la innovación o creación tecnológica que de alguna forma contribuye para el mejoramiento de la vida o de una comunidad. Así, para que realmente las soluciones derivadas de estos proyectos de estudios sean aplicadas y llevadas en cuenta por el grupo de gestores, coordinadores o líderes de algún servicio específico, se hace necesario un especial cuidado con la parte de presentación y comunicación de los hallazgos y/o soluciones del proyecto en cuestión.

El capítulo que sigue marca el inicio de la discusión con respecto a la relación entre la manera de funcionar de un laboratorio social con enfoque en ciencias de datos y determinados aspectos observado en la sociedad occidental contemporánea, por parte de los estudios sociales de investigadores interesados en la temática. Observar este contexto de investigación nos permite encontrar hallazgos sobre la manera como la sociedad viene siendo representada en las ciencias sociales latinoamericana, a partir de sus publicaciones académicas (estudio bibliométrico Scopus y Scielo) y las actividades científicas a través de la introducción de las técnicas big data en contextos de investigación en ciencia de datos.

### **Cap 9: (Posibles) Implicaciones sociales de una sociedad computacional**

Como se ha discutido, sabemos que Karin Knorr-Cetina (2005) ha argumentado que los laboratorios de investigación son espacios cruciales para comprender cómo se generan y validan los conocimientos en la sociedad. Ella ha enfatizado la importancia de analizar no sólo el producto final de la investigación (los textos científicos), sino también los procesos, las prácticas y las interacciones que ocurren dentro de los laboratorios. Esta comprensión introduce no solamente el imperativo por la búsqueda de las dinámicas y negociaciones de diversas naturalezas pertenecientes a los laboratorios, también contribuye para el entendimiento de la dimensión relacional existente entre el contexto social, las prácticas de la sociedad y las prácticas de laboratorio (Knorr-Cetina, 2005).

La relación entre los estudios de laboratorio y la sociedad, según la visión de la autora, es fundamentalmente una relación de afectación recíproca. Ha destacado la importancia de comprender cómo los laboratorios de investigación y las prácticas científicas influyen y son influidos por la sociedad en la que operan de más de una forma distinta.

El primer factor que respalda su visión es el hecho de reconocer la contextualización del conocimiento. Lo que significa que el conocimiento científico no se produce en un vacío, sino que está fuertemente influenciado por el contexto social, cultural y tecnológico en el que se encuentra el laboratorio. Los científicos, como parte integrante y esencial de estos procesos no son actores aislados y tampoco neutrales. Siendo humanos, operan sus actividades y prácticas laboratoriales a partir de normas y valores, intrínsecamente relacionados a normas y valores sociales circunscritos en un tiempo y un espacio particular (Knorr-Cetina, 2005).

Otro argumento sólido, que demuestra la relación de interconectividad entre sociedad y producción del conocimiento científico, es el argumento del conocimiento situado. Este está arraigado en prácticas específicas y en un entorno particular. Los laboratorios de investigación son lugares donde se produce este conocimiento situado, y la forma en que se llevan a cabo las investigaciones está moldeada por las condiciones y contextos de la sociedad circundante (Knorr-Cetina, 2005).

Por otro lado, Knorr-Cetina (2005) también señala que los avances científicos y las investigaciones realizadas en laboratorios tienen un impacto significativo en la sociedad. Los resultados de la investigación pueden influir en políticas públicas, en la toma de decisiones y en la percepción pública de temas científicos, agendas nacionales e incluso globales.

La translocalidad científica<sup>110</sup> también es una realidad que impacta la vida social: diversos autores han estudiado cómo la ciencia se ha vuelto cada vez más global (Feld y Kreimer, 2019) y conectada a través de redes internacionales (Albornoz, 2002), el laboratorio se vuelve un lugar capaz de expandir sus ramificaciones globales, influenciando y afectando a comunidades y sociedades en todo el mundo (Demarchi, 2020).

En definitiva, los laboratorios de investigación científica son espacios donde la sociedad y la ciencia interactúan de manera compleja y recíproca. Siendo el laboratorio, el lugar del trabajo del hombre y “producto de relaciones de fuerza, de negociaciones (y, deberíamos agregar, de

---

<sup>110</sup> La translocalidad científica se refiere a la tendencia de la investigación científica y de la actividad académica a trascender las fronteras nacionales y a ocurrir en un entorno globalizado. Esto implica la colaboración de científicos y académicos de diferentes países en proyectos de investigación comunes, la circulación global de conocimientos y la interconexión de la comunidad científica a nivel internacional. La translocalidad científica se caracteriza por la movilidad de investigadores, la participación en redes internacionales y la capacidad de abordar problemas y desafíos globales a través de la colaboración internacional.

sumisiones y de imposiciones” (Knorr-Cetina, 2005, p.37), también es, por consecuencia, socialmente construido.

De la derivación de este raciocinio es que se apoya el entendimiento de que la sociedad influye en la ciencia a través de sus valores, normas y expectativas, mientras que la ciencia, a su vez, influye en la sociedad a través de sus descubrimientos, tecnologías y aplicaciones. La comprensión del papel fundamental de la interacción entre laboratorios y sociedad es crucial para entender la manera en que Knorr-Cetina concibe el constructivismo en sus estudios de la ciencia<sup>111</sup>.

Consciente de estos atributos, y de esta manera de comprender la fabricación de cualquier tipo de conocimiento, afirmo que esta presente investigación, se constituye como un documento que registra el traspaso de un concepto, como fue demostrado: desde big data a las ciencias de datos y la madurez de una profesión (el especialista social en (ciencia) datos). Este trabajo de investigación también es testigo de las transformaciones y la introducción de nuevos modos de abordar la manera de ser sociedad y estudiar la sociedad a través de la ciencia.

Es por esto que arriesgo declarar que los hallazgos encontrados a lo largo de esta investigación, analizados bajo la óptica de mi propias selecciones, posibilidades de observación y recursos disponibles, implican también una serie de repercusiones, o comprensiones con respecto a cambios e implementaciones de actividades prácticas, tanto en la propia ciencia, como en la sociedad. Estamos ejerciendo nuestro oficio de científicos sociales y observando el desarrollo de la sociedad al mismo tiempo en que somos afectados por las transformaciones tecnológicas que hemos experimentado en los últimos 10 años.

Un claro ejemplo de lo anterior, fue constatar cómo cambió el uso del término “big data” a “ciencias de datos” en el universo investigativo no solamente en las ciencias sociales. También lo percibimos en otras áreas del conocimiento en latinoamérica en el mundo. Del mismo modo, puede afirmarse que la sociedad también experimenta este proceso de madurez al incorporar vocablos y conceptos tecnológicos en su vida cotidiana: las personas se

---

<sup>111</sup>En la introducción a la versión traducida del libro de Knorr-Cetina de 1981, Pablo Kreimer escribe una brillante reseña con respecto a la visión constructivista de los estudios de la ciencia, tecnología y sociedad, donde apunta las diferencias entre los estudios de laboratorio de Latour y Woogar (1986), y Lynch (1993), en comparación a de Knorr-Cetina. Introduce conceptos metodológicos y etnográficos que la diferencian de los demás célebres nombres de este campo de estudio, postura que asumí como base teórica para la construcción de esta investigación.

interesan y entienden más sobre estos temas, se familiarizan con herramientas computacionales, sistemas y aplicaciones basadas en la computación, inteligencia artificial y la automatización. Basta estar atento a los diálogos en la calle para notar que palabras como “algoritmo”, “IA” y “programación” están siendo utilizadas cotidianamente como parte intrínseca a la vida contemporánea.

No faltan ejemplos que conduzcan a la discusión sobre el impacto de las nuevas tecnologías y el mercado de compras en internet (Toledo, 2020; Antunes y Maia, 2018), ni la mirada con respecto a las decisiones de compra, adoctrinamiento algoritmo de la vida, comportamiento del consumidor, entre otras.

Otro ejemplo, las implicaciones del uso de las herramientas y aplicaciones tecnológicas de movilidad urbana y geoespacial, ha condicionado cada vez más la manera de vivir y experimentar la ciudad (Ruiz, 2020; Arma-Beatriz, Reis-Martins y Carranza-Infante, 2020). Los mecanismos de control del espacio y seguridad pública han cambiado la manera en la que nos relacionamos con el espacio, y ha introducido en nuestro vocabulario digital términos como *reconocimiento facial*, *algoritmo sospechoso*, entre otros (Villalobos, 2020). Hemos adaptado la manera de educar a las nuevas generaciones en función del interés y de procesos de inclusión digital, lo que ha contribuido a la creación de nuevos e innovadores espacios de aprendizaje y metodologías de enseñanza (Mendoza, 2018; Amaya-Amay et al., 2020).

La forma de medir la percepción pública también ha cambiado en función de las redes sociales, o maneras alternativas de recolectar la opinión de las personas (Vinícius y Calda, 2021). En el área de derecho han surgido importantes debates, como la discusión que se ha venido ampliando en la sociedad sobre los riesgos de violación de derechos a la protección de datos y anonimidad (Benevides, 2017), por dar algunos ejemplos.

Además de que el trabajo de los científico abarca más allá de trabajar con datos masivos, el cambio de término tiene que ver con la percepción nítida de la velocidad y el alto grado de profundización en las dinámica de la sociedad contemporánea, que están inmersa en la era de la digitalización y la información. Esta transición no sólo es un ajuste conceptual, sino un reflejo de cómo la ciencia de datos, la ciencia computacional y su aplicación en las ciencias sociales están en sintonía con la creciente importancia de la tecnología y los datos en la vida cotidiana, afectando la toma de decisiones, la política, la economía, la vida social y la forma en que interactuamos en la sociedad moderna.

Para captar y evidenciar estas transformaciones mencionadas por esta investigación, con el objetivo de dar más cuerpo y sustentación a los hallazgos que se presentará en la secuencia, utilice de la conjunción de los resultados obtenidos través de mi trabajo de campo etnográfico y el proceso de análisis de las narrativas generadas por medio de las entrevistas realizadas con algunos de los autores de las publicaciones analizadas a partir del estudio bibliométrico realizado. Así, las entrevistas fueron utilizadas como fuente de información para completar el estudio de las publicaciones académicas.

A medida que avanzaba el estudio de campo, quedó claro que muchos aspectos observados en las prácticas, resultados de investigación y análisis de proyectos del GobLab, laboratorio del estudio etnográfico, resonaban con los hallazgos e interpretaciones de una gran parte de los textos revisados en mi base de datos bibliométrica. Para corroborar y profundizar en estas coincidencias, surgió la necesidad de dialogar con el mayor número posible de autores de esas publicaciones identificadas en la fase bibliométrica. Tras contactarlos y confirmar su interés, la mayoría de las entrevistas se realizaron en plataformas de reunión en línea, siguiendo un guión previamente preparado (ver anexo). Entonces, concreté citas on line para debatir con estos autores sobre sus textos, el contexto de producción de su investigación, interés, trayectoria académica y su proximidad con el tema de los datos masivos.

Considero que las entrevistas semi-estructuradas a los autores y científicos del material investigado, son parte de la trayectoria de investigación como un todo (Díaz y Navarro, 1995). La importancia de las entrevistas está en que desvelan aspectos que van más allá de los “indicadores objetivos” de la investigación estratégicamente elegidos para estar presente en sus publicaciones. Es mediante la conversación y el análisis más atentos a los discursos conducidos por los científicos que se torna posible dar a conocer los procesos de selectividad de sus elecciones, la artesanía y el camino tomado para su quehacer.

Ese contraste no sólo se verifica entre la fragmentariedad y la inmediatez del discurso del laboratorio y la coherencia editada y pulida del discurso escrito. En el laboratorio, el razonamiento científico exhibe sus preocupaciones con pureza salvaje. Pero en sus artículos, los razonadores salvajes del laboratorio parecen cambiar de fe. El razonamiento del artículo, se puede pensar, es fiel a las *scripturas* (las escrituras de autoridad) de un área, más que a las preocupaciones a partir de las cuales se originó. Pero al mismo tiempo ese razonamiento contiene una conspiración para apropiarse o destituir parte de esas *scripturas*. En su superficie civilizada y mansa, el artículo científico esconde más de lo que dice. Por una razón: se olvida deliberadamente de

mucho de lo que ocurrió en el laboratorio, aunque pretende presentar un "informe" de esa investigación. Segundo, los productos escritos de la investigación emplean una buena cantidad de estrategias literarias que en gran medida pasan desapercibidas para los lectores. (Knorr-Cetina, 2005, p. 224)

Sabemos que el artículo es un proceso de negociación entre autores y críticos, demostrando también una preocupación no solamente con la técnica, sino también con el control social por detrás de las selecciones de contenidos presentes o no en el texto académico. Es, por lo tanto, una tentativa a un acercamiento a las historias por detrás de la investigación: otras evidencias de lo que se pasa en los bastidores de la investigación, en un territorio llamado laboratorio. Como diría la misma autora: “La escritura del artículo comienza mucho antes de que esté escrito el borrador, mediante las huellas que se han ido generando durante el trabajo de laboratorio” (Knorr-Cetina, 2005, p. 247).

Lo que nos hace volver al punto original epistemológico que conlleva la presente investigación: considerar la investigación científica como constructiva y no como descriptiva, ver los productos científicos como una elaboración intencionada internamente en términos de la selectividad que incorporan. Estudiar la investigación científica, más allá del texto, es, entonces, estudiar el proceso por el cual se efectúan las respectivas selecciones.

Para realizar tales entrevistas, busqué la gran mayoría de los autores de las publicaciones de la base de datos Scielo primeramente. Curiosamente obtuve la respuesta apenas de los artículos más recientes, publicados después de 2017. Entrevistando así autores de publicaciones de 2020, una de 2019 y una de 2018.

Con respecto a la base de datos Scopus, el número de conversaciones fue más numeroso debido a mayor expresividad de textos provenientes de esta base de datos. Más allá de los 17 autores entrevistados de publicaciones en las Ciencias Sociales, en la ocasión de la visita de Henry Chávez a Chile, se ejecutan siete entrevistas colectivas (ambos desempeñando papel de entrevistadores simultáneamente) con autores de la misma librería y no necesariamente vinculados a las ciencias sociales. Durante este periodo entrevistamos a profesionales y académicos de las demás áreas del conocimiento, que también habían escrito textos sobre big data.

También tuve acceso, gracias a la colaboración de Henry, a otras 20 entrevistas que él realizó durante sus visitas de investigación postdoctoral en Ecuador y México. Finalmente, durante mis visitas a Brasil a congresos y similares, tuve la oportunidad de entrevistar personalmente a 9 personas, entre autores de artículos y recomendaciones de estos mismos, totalizando así un número de 61 entrevistas analizadas, datos que sintetizo a continuación. Para leer el guión, dirigirse al anexo.

**Tabla 31.** *Relación de entrevistas realizadas y analizadas para la investigación*

	Autores Scielo Ciencias Sociales	Autores Scopus Ciencias Sociales	Colectiva Presencial Chile	Compartidas por Henry: Ecuador + México	Autores y Info Claves Brasil	total de entrevistas accedidas
Cantidad de Entrevistas	8	17	7	20	9	61

Es, por lo tanto, a partir del movimiento de mi propia construcción de conocimiento, que presento los capítulos subsecuentes, sin olvidar el hecho de que:

Evidentemente, sólo el tipo de material aún no resumido ofrecido por los grabadores y los videograbadores puede ser sometido a nivel de análisis de microprocesos que quisiéramos que acompañe a una metodología sensitiva. Sin embargo, aun cuando ese material tiene la ventaja de no estar resumido, *no* está no construido: las técnicas de transcripción y categorización del comportamiento no verbal, lo parcial de la grabación tomada por una cámara o el cambio de comportamiento provocado por la presencia de un grabador, todos apuntan en dirección a la selectividad incorporada en ese material. Como en el caso de los productos científicos, los resultados de las ciencias sociales tienen que ser vistos también como selectivamente contruidos. El punto de una sociología sensitiva no es remediar la constructividad sino para pedir prestado ese término introducido páginas atrás, *descentrar la constructividad* de modo tal de que pase a ser una empresa intersubjetiva. Que tenemos que tomar grandes recaudos para permitir que el campo del estudio ejerza realmente los constreñimientos deseados en la información interpretada es algo que ha sido demostrado por el desarrollo de la antropología, que hace largo tiempo denunció el etnocentrismo societal, aunque eso no la ha librado de seguir inmersa en su propio etnocentrismo profesional. (Knorr-Cetina, 2005, p.92)

Una verdadera metodología sensitiva no puede ignorar su dinámica, ni el carácter temporal de su propia investigación. Está claro que las unidades de observación y explicación deben permitir que emerjan los rasgos temporales, contextuales e interaccionales de la acción, pues

es un retrato temporal y demarcado por un contexto que expresa interpretaciones tanto a nivel personal (autores, laboratorio y la mía), como también colectivos.

Con este enfoque aclarado y expuesto el modelo de producción y recolección de información que guiará la presentación de hallazgos, se procederá a mostrar los resultados de esta combinación metodológica. Este enfoque integra el análisis de contenido de los textos publicados y revisados en Scielo y Scopus, los hallazgos de la fase etnográfica de la investigación y el análisis de las narrativas derivadas de las entrevistas realizadas a 61 autores de dichas publicaciones.

### **9.1 La sociedad computacional: representación de lo social en los estudios de ciencias sociales**

El uso de técnicas computacionales, en el contexto de la ciencia de datos, para análisis de información social, observada en el laboratorio etnografiado, en las entrevistas emprendidas con académicos especializados y los autores de los textos aprendidos, en conjunto, demuestran un entendimiento de algunas características común, similar o frecuentes que apuntan a la representación de una sociedad a la cual he propuesto llamar *sociedad computacional*.

Lo que llamo de *sociedad computacional* es un patrón de comportamiento de la sociedad contemporánea presentado por los científicos sociales en la conducción de sus estudios e investigaciones. Se constituye como un apañado de las características más frecuentes presente en el discurso proveniente de los trabajos investigativos de las publicaciones y narrativas construidas alrededor de las entrevistas realizadas. Lo que se hizo fue condensar en siete principales características, las constataciones más comunes, más frecuentes presente en dichos trabajos.

La *sociedad computacional*, es por lo tanto el resumen, una especie de fotografía que está empezando a ser revelada, por medios de los estudios de los científicos sociales y sus prácticas laborales al interior de sus proyectos de investigación y los hallazgos del trabajo por mí conducido. Las dimensiones encontradas: la conectividad, el tecno-encanto, la automatización de la vida, la velocidad, la identidad global, la alienación del trabajo por

detrás de la máquina y la identidad anfibia, se constituyen como patrones comportamentales de una sociedad cada vez más expuesta a la computacionalización de la vida.

Como ya he descrito a lo largo de este texto, la tendencia de la digitalización de la vida social, que entra con fuerza a partir de los años 2010, se ha incorporado a la práctica investigativa y a la producción del conocimiento de las ciencias sociales (Ponce-Tituaña y Lucio-Paredes, 2021). La acelerada digitalización de la vida social, no es simplemente un fenómeno transitorio; más bien, marca la consolidación de una sociedad cada vez más computacional.

La *sociedad computacional*, por lo tanto, está constituida como un hallazgo de esta investigación, y están conformadas a partir de siete dimensiones que consolidan el discurso actual en las ciencias sociales latinoamericanas interesadas en la relación entre sociedad, tecnología computacional y datos masivos.

Hablar de una sociedad computacional, en este contexto, no sólo se refiere a la omnipresencia de la tecnología computacional en nuestras vidas diarias, sino también a la manera en que las ciencias sociales latinoamericanas, a través de su trabajo investigativo, viene priorizando la forma cómo representa, concibe y estudia la sociedad contemporánea. Estos imperativos, entonces, se constituyen como la base que moldea nuestras interacciones, instituciones y formas de conocimiento, que viene recién instaurándose.

A título de resumen y mejor visualización de tales resultados, antes de adentrarse con más profundidad en cada una de ellas, presento una tabla que sintetiza estas dimensiones encontradas con mayor frecuencia en mis análisis:

**Tabla 32.** *Paradigmas de la ciencia de datos y la sociedad computacional*

Característica	Producción del Conocimiento (Ciencia)	Sociedad Contemporánea (Sociedad)	Característica
<b>Disminución de los tiempos de Investigación</b>	Aceleración de los procesos de obtención de resultados a través del aprovechamiento de técnicas de procesamiento computacional y análisis avanzada de grandes de datos	<p>Aceleración de la comunicación, el acceso a información y a la toma de decisiones en tiempo real, provocando cambios en la interacción y expectativas de respuesta a diversos estímulos sociales.</p> <p><b>DIMENSIONES ASOCIADAS:</b></p> <ul style="list-style-type: none"> <li>● Rapidez en la comunicación</li> <li>● Acceso inmediato a la información global</li> <li>● Retroalimentación en tiempo real</li> <li>● Expectativa de urgencia</li> </ul>	<b>Instantaneidad de la vida</b>

		<ul style="list-style-type: none"> <li>• Decisiones precipitadas</li> </ul>	
<b>Abaratamiento de la Investigación</b>	Uso de instrumentos tecnológicas que reducen los costos asociados a la Investigación recolección y análisis de datos, haciendo que la investigación sea más accesible económicamente.	<p>Reducción de costos en la provisión de servicios en consecuencia de los avances tecnológicos.</p> <p>DIMENSIONES ASOCIADAS:</p> <ul style="list-style-type: none"> <li>• Automatización y eficiencia</li> <li>• Economías de escala</li> <li>• Mayor competitividad</li> <li>• Reducción de intermediarios</li> <li>• Personalización en la entrega de servicios</li> <li>• Teletrabajo y colaboración en línea</li> </ul>	<b>Abaratamiento de los costos de servicios</b>
<b>Principio de la Eficiencia</b>	Enfoque en la optimización de los recursos utilizados para obtener resultados más rápidos y precisos.	<p>Búsqueda por mayor productividad y eficiencia de la vida a través de la apropiación de herramientas tecnológicas en los ejercicios y actividades cotidianas.</p> <p>DIMENSIONES ASOCIADAS:</p> <ul style="list-style-type: none"> <li>• Acceso a recursos globales</li> <li>• Personalización de servicios</li> <li>• Automatización - reducción de errores humanos</li> <li>• Colaboración en línea</li> </ul>	<b>Principio de la Productividad</b>
<b>Alto grado de experimentación</b>	Capacidad de manipular y analizar grandes conjuntos de datos de manera ágil, lo que permite realizar experimentos y pruebas a gran escala.	<p>Habilidad para ajustarse rápidamente a nuevas tecnologías y a los cambios que requieren su utilización, en un mundo en constante evolución tecnológica.</p> <p>DIMENSIONES ASOCIADAS:</p> <ul style="list-style-type: none"> <li>• Rapidez en la adopción tecnológica</li> <li>• Cultura de aprendizaje continuo</li> <li>• Flexibilidad en el lugar de trabajo</li> <li>• Resiliencia ante cambios tecnológicos</li> <li>• Interconexión global</li> </ul>	<b>Alto Grado de Adaptación</b>
<b>Innovación</b>	<p>Imprevisibilidad, error y pruebas como catalizadores de la innovación.</p> <p>Contexto que incentiva los enfoques novedosos al abordar preguntas de investigación.</p>	<p>Resultado de la innovación y la constante aparición de nuevas tecnologías, aplicaciones y conceptos en la sociedad actual.</p> <p>DIMENSIONES ASOCIADAS:</p> <ul style="list-style-type: none"> <li>• Ampliación del espectro de la toma de decisión</li> <li>• Diversidad de plataformas, dispositivos y soluciones</li> <li>• Cultura de la personalización</li> <li>• Inclusión y representación</li> </ul>	<b>Variedad - Diversidad</b>
<b>Gran conjunto de datos</b>	Grandes y diversas cantidades de datos como recurso en la obtención de información detallada y valiosa.	<p>Ampliación la accesibilidad global, disminución de las fronteras geográfica y interconexión en términos culturales, económicos, sociales, políticos y científicos</p> <p>DIMENSIONES ASOCIADAS:</p> <ul style="list-style-type: none"> <li>• Acceso facilitado a información y recursos globales y en línea</li> <li>• Conexión y cooperación internacional</li> <li>• Uniformidad de la vida y peculiaridades culturales</li> <li>• Comunicación de emergencia</li> </ul>	<b>Accesibilidad Global</b>
<b>Preponderancia cuantitativa</b>	Énfasis en el análisis cuantitativo de datos para respaldar hallazgos y tomar decisiones basadas en evidencia.	<p>Valorización de la vida a través de aspectos cuantitativos. Permitiendo brindar un sentido de logro, a la vez que plantea desafíos relacionados al desarrollo de una noción de satisfacción más profunda.</p> <p>DIMENSIONES ASOCIADAS:</p> <ul style="list-style-type: none"> <li>• Énfasis en el progreso material</li> </ul>	<b>Calificación Cuantitativa de la vida</b>

		<ul style="list-style-type: none"> <li>• Medición del éxito</li> <li>• Presión por el consumo</li> </ul>	
Consenso general	Mejores investigaciones	Mejor bienestar	Consenso general

Como se puede ver, busco establecer un paralelo entre las principales tendencias observadas en un laboratorio social en ciencias de datos, y los estudios interesados en la temática y los distintos aspectos del comportamiento social contemporáneo, frecuentemente señalados en las publicaciones analizadas en la fase de estudio bibliométrico. Para ello, me basé en los textos analizados y en algunas de las entrevistas realizadas para esta investigación. De este modo, delimito la representación social de una sociedad emergente, propuesta por estos trabajos. Desde mi perspectiva, estos principios definen la identidad de la sociedad que estamos construyendo a medida que incorporamos tecnologías de procesamiento computacional de datos a otras esferas de la vida social contemporánea, según la visión construida a partir de los trabajos de los científicos sociales, estudiados en esta investigación.

Estas siete características o principios, buscan mostrar la manera en que la sociedad actual es representada y concebida por la literatura académica actual, considerando mi muestra bibliométrica y la bibliografía estudiada. También forman parte de la comprensión reflexiva que pude concluir, a través de los cruces de las diversas técnicas utilizadas en el proceso de recolección de información de esta investigación.

Antes de seguir, es importante comentar que aunque estos principios están separados en forma de subtópicos, a título de mejor ordenamiento del texto, en la práctica, están íntimamente conectados e interrelacionados. El ordenamiento por tópicos fue una decisión para organizar mejor las ideas contenidas alrededor de su concepto.

A continuación, analizaré la relación entre las características observadas en contextos de investigación junto a su contraparte en la sociedad, basándome en mis observaciones y en los paradigmas e interpretaciones derivadas de los enfoques constructivistas de los estudios de ciencia, tecnología y sociedad presente en las narrativas de las entrevistas realizadas con algunos de los autores de las diversas publicaciones analizadas en el fase bibliométrica de investigación..

Profundizaré en las dinámicas de la sociedad computacional a partir de la elucidación de dichos principios, examinando cómo la computacionalización de la vida no sólo transforma nuestras actividades cotidianas, sino que da forma a una nueva narrativa societal donde la conectividad y las tecnologías computacionales digitales son elementos centrales en la construcción de nuestro presente y futuro colectivo.

### ***9.1.1 Conectividad***

El imperativo de la conectividad, se refiere a la creciente necesidad de asegurar y posibilitar estructuralmente, que la mayor cantidad de personas posible tenga acceso a Internet y esté conectada en la sociedad digital actual. La conexión se ha convertido en una exigencia fundamental de la vida moderna, ya que, en numerosas ocasiones, es la única manera en que se puede acceder a información, oportunidades y servicios; de participar en la economía y en las actividades de la sociedad en general.

Ello y nuestra múltiple convivencia en espacios digitales, como consecuencia, genera la capacidad de producir cada vez más datos, que a la vez, retroalimentan la necesidad de mayor cobertura y expansión de la conectividad de la vida. Es como un círculo vicioso: más conectividad, más datos, más datos digitales; luego, mayor la necesidad de garantizar y expandir la conexión global. La conectividad de las personas, de los servicios y de la vida social, empuja y alimenta la necesidad de obtener más datos. A su vez, estos necesitan de un ambiente de más y más conectividad para mejorar y expandirse. De la misma manera, el conocimiento computacional y la utilización de sus técnicas para el procesamiento de grandes datos, ha ido creciendo y ganando espacio a medida que la conectividad mundial también crece.

La sociedad computacional no sólo es testigo de una mayor dependencia de la tecnología, sino que representa un cambio en la misma estructura de nuestra vida social, donde la conectividad no es sólo una conveniencia, sino un imperativo que influye en la configuración de nuestras identidades, relaciones y comunidades.

En este contexto, las tecnologías digitales no sólo sirven como herramientas para facilitar el trabajo remoto, la educación en línea y otras actividades cotidianas, sino que se han integrado en la esencia misma de la vida social y, por supuesto, en la producción de conocimiento. La

ciencia social, como hemos explorado en este texto, ha experimentado una transformación paradigmática, adoptando metodologías digitales y adaptándose a un entorno donde la conectividad instantánea y la colaboración virtual son una de las prerrogativas.

A medida que exploramos las implicaciones de esta sociedad computacional, es crucial analizar no sólo los aspectos prácticos de la tecnología, sino también su impacto en la forma en que construimos el conocimiento y en que entendemos nuestra realidad social. Este análisis se sitúa en el cruce de la tecnología, la sociedad y la academia, delineando un terreno fértil para la investigación que busca comprender y abordar los desafíos y oportunidades de esta nueva era digital.

La incorporación de las Tecnologías de la Información y Comunicación (TIC), junto con la importancia de la conectividad y el encanto hacia la tecnología, tienen un impacto significativo en la forma en que las sociedades conviven. Estos cambios influyen en nuestras acciones, expectativas sobre la vida y en los ideales de comportamiento que cada sociedad persigue.

¡Hay que conectar a la gente! no podemos olvidar esto. Y aunque tenemos la impresión de que todos tienen acceso a internet, a un celular con 3G, o a un computador, falta mucho todavía para que el mundo esté en línea. Y esto es importante no solamente por una cuestión de democratización de la digitalidad o de las tecnologías en general, también porque es una necesidad mundial. Hablo de empresas y de gobiernos. No es solamente necesario estar conectado para comunicarse con la familia de cualquier parte y “sin costos”, o, que sé yo... buscar a tu pareja en un app de relaciones, o para saber moverse por la ciudad. Estar conectado es importante porque incluso los gobiernos están cada vez más digitales. Yo diría que en 5 años más, si una persona no tiene un correo electrónico, y no tiene acceso a internet no va poder recibir un beneficio del Estado, en caso de que lo necesite, o no va poder participar de un concurso público, no va poder registrarse ¡u obtener la cédula de su hijo que acaba de nacer! Entonces, conectar a la gente pasó de ser lujo a ser una necesidad básica en cuando un ciudadano del mundo de hoy. ¿Es triste? no sé, pero es la realidad que viene imponiéndose. (JLP - Entrevista 7/2022 - Brasil - mi traducción)

Del mismo modo en que observamos que el sistema de investigación en ciencias sociales refleja las características de una sociedad que tanto influye como es influenciada por las mismas dinámicas que rigen un laboratorio de innovación pública en una universidad chilena. Sostengo que el estudio etnográfico realizado en diversos proyectos de investigación social a

partir de la utilización de técnicas de procesamiento de grandes datos, sugiere que principios similares también emergen en el tejido social.

Abajo, una tabla esquemática, que resume las ideas ideas generales de la manera como este principio es representado en los estudios de ciencias sociales, en contexto de investigación en ciencia de datos:

**Tabla 33.** *Resumen esquemático conectividad*

Conectividad	Palabras Asociadas	Patrones Instaurados
	<ul style="list-style-type: none"> <li>* Conexión</li> <li>* Internet</li> <li>* Vida Moderna</li> </ul>	<ul style="list-style-type: none"> <li>- Todo el mundo en un solo lugar</li> <li>- Sociedad Digital</li> <li>- Accesibilidad</li> <li>- Participación</li> <li>- Información Ilimitada</li> <li>- Dependencia Digital</li> </ul>

### 9.1.2 *Tecno-encanto*

El tecno-encanto es un principio que tiene que ver con el encantamiento, la fascinación y, a veces, la sobrevaloración del papel de la tecnología en la sociedad contemporánea. Bien lo ilustra el escritor Vivek Wadhwa:<sup>112</sup> “La tecnología abrirá camino a la iluminación”. Vivimos en una época donde la tecnología no sólo se ha vuelto omnipresente, sino que también ha capturado nuestra imaginación y ha afectado la forma en que percibimos el mundo que nos rodea. Este principio, el principio del Techno Encanto, refleja la predisposición inherente de la sociedad computacional a encontrar magia y maravilla en los avances tecnológicos, generando una confianza inquebrantable en que, a través de la continua evolución tecnológica, el futuro se teje gracias a esta idea a partir de un imaginario colectivo de progreso y mejora constante.

Hay una creencia de que la solución a los desafíos humanos y sociales residen en el desarrollo tecnológico, a la vez que vemos una fascinación por esta innovación tecnológica al

---

<sup>112</sup>Nacido en la India, Vivek Wadhwa se graduó en la Universidad de Canberra (Australia) en Ciencias de la Computación. MBA en la Universidad de Nueva York. Es autor de *The Immigrant Exodus: Why America Is Losing the Global Race to Capture Entrepreneurial Talent*, el mejor libro de 2012, según The Economist. También escribió *Innovating Women: The Changing Face of Technology*. (2014) junto con Farai Chideya. Dicta clases de entrepreneurship y políticas públicas en las universidades de Stanford y Duke. Supervisor de investigación en la Singularity University.

punto que define nuestras interacciones diarias, repercutiendo en la toma de decisiones a nivel individual y colectivo, permeando cada aspecto de nuestra vida en la sociedad computacional.

El principio tecno-encanto, por lo tanto, se define como la fascinación o entusiasmo social y colectivo que se experimenta en la sociedad contemporánea debido a los avances tecnológicos y las tecnologías digitales. Es la sensación de que la tecnología actual es un símbolo de progreso, desarrollo, innovación, velocidad, mejores decisiones y una mayor comprensión del mundo. El tecno encanto refleja la admiración generalizada y el optimismo en torno a las posibilidades que ofrecen las tecnologías digitales para mejorar la vida y la sociedad, así como el reconocimiento de su creciente influencia en diversas áreas de la vida cotidiana.

La fascinación por la tecnología, está totalmente ligada a la creencia del avance de la ciencia, de las técnicas y de la capacidad del hombre de dominar la máquina, aprovechando mejor sus recursos. Si bien hay recelo junto a la expectativa sobre las desventajas del avance tecnológico, un cierto miedo sobre si las máquinas dominarán al hombre, en general se percibe un tono positivo e incluso mesianico con respecto al avance de las tecnologías y el universo de nuevas oportunidades y facilidades para la vida que a los humanos pueden traer.

La discusión sobre los riesgos del avance tecnológico también alcanza los muros de las universidades, laboratorios y espacios académicos. Con los científicos que tuve contacto durante esta investigación, surge en la conversación los riesgos y sesgos generados por los algoritmos, como racismo y clasismo. También se habla sobre la amenaza de desempleo frente a diversos oficios humanos siendo remplazados por máquinas. Sin embargo, las publicaciones estudiadas evidencian que todavía estamos en una fase de asombro por la tecnología, donde cualquier novedad emergente, una vez comprendida su función, despierta un profundo sentido de admiración por parte de sus usuarios, sean ellos ciudadanos en su cotidianidad laboral y de vida, o académicos en el ejercicio de sus funciones investigativas.

Los artículos presentados en los antecedentes de la tesis ilustran y evidencian este entusiasmo hacia la innovación en metodologías y enfoques de investigación, a pesar de que nosotros, los científicos sociales, seamos quienes más advierten sobre los riesgos de la introducción de estos avances tecnológicos en nuestra sociedad..

Este optimismo se expresa muy bien en lo que compete a este estudio: la posibilidad de conectar los datos a través de los recursos tecnológicos y digitales, han proporcionado una mayor conectividad de los métodos, los sujetos de investigaciones, los sistemas sociotécnicos a su alrededor y la fuerza de trabajo necesaria para llevar a cabo la actividad de investigación. La digitalización y la conectividad de la vida genera mayor información, lo que, a su vez, genera una mejor capacidad de transformar computacionalmente esta información disponible en soluciones y en resultados relevantes. Recordemos que, mientras más (y buenos) datos, mejor potencial de análisis dentro del paradigma de producción de conocimiento en investigaciones sociales a partir de técnicas de procesamiento computacional.

Es bueno resaltar, con todo, que a pesar de los avances tecnológicos y el esfuerzo de conectar el globo a través de internet sea innegable, aún existe una brecha digital significativa en todo el mundo. Lo que implica aún más cautela con respecto a las inferencias y conclusiones que pueden ser levantadas en cuanto a los estudios, principalmente sociales, de la sociedad en general. En lo que dice respecto específicamente al contexto de las investigación, los especialistas en humanidades con los cuales hemos conversado a través de las entrevistas, son categóricos en reconocer que las investigaciones sociales hechas a partir de, por ejemplo, redes sociales y opinión pública, presentan limitaciones importantes en cuanto a la brecha digital mundial y el corte de clase social.

Por otro lado, es posible percibir una gran esperanza por parte de estos mismos científicos, que apuestan que “el pasar del tiempo” permitirá que la capacidad de representatividad de la información obtenida a través del procesamiento computacional de los datos digitales sea tan o más rigurosa que los parámetros estadísticos tradicionalmente utilizados:

Yo no me preocupo tanto con esto porque es cuestión de tiempo [hablando de la dicotomía entre la investigación ser o no ser muestral]. En muy poco tiempo vamos a ver todo el globo conectado a internet vía satélite. Y esto no es una especulación, estamos viendo el masivo lanzamiento de satélites de Space X, para conectar el mundo. Esto ya es el presente! Entonces yo creo que en poco tiempo vamos lograr mejorar nuestra capacidad de incluso revolucionar los métodos estadísticos. Cuando llegamos a este nivel, ni siquiera nos vamos a cuestionar más sobre la solidez de un estudio porque no es tradicionalmente amostral. Será. Pero de otra manera. Será cada vez más fácil, tener los datos de cada una de las personas, así que la muestra en sí no será el problema. La gran cosa es saber cómo vamos crear estas nuevas reglas de amostragem, más adaptadas a la realidad del mundo actual. (HRZ - Conversación - 2022 - Chile)

El habla de este científico refleja un tono de optimismo en relación al futuro y el devenir de las ciencias, una confianza y compromiso con lo que está por venir, lo que está estrechamente vinculado a las prerrogativas generadas alrededor del concepto de tecno encanto, experimentada en el presente.

Es por eso, también, que estamos viendo cada vez más iniciativas gubernamentales a favor de la inclusión digital en sus países. Muchos gobiernos están reconociendo la importancia de reducir la brecha digital y están implementando iniciativas para proporcionar acceso a Internet a comunidades desatendidas. Estas iniciativas pueden incluir la expansión de infraestructura de banda ancha, la creación de zonas de acceso público a Internet y programas de subsidios para la adquisición de dispositivos.

Se ve, además, un claro esfuerzo por levantar programas de digitalización del Estado, desde documentos a programas internos, como es el caso del sistema Bem-te-vi, el proyecto Ulysses, el asistente virtual de la DPP, el programa de detección de fraude de FONASA. Otra área en creciente expansión en función del esfuerzo de digitalizar el Estado, es lo que lo llaman de “relación con el cliente”, o sea, los servicios que dialogan directamente con los ciudadanos, como por ejemplo la designación de beneficios, emisión de documentos, trámites civiles, evaluación de programas y políticas, canal de atención ciudadana. Estos recursos están cada vez más asociados con la capacidad y operación del gobierno digital y la inclusión de la población a estos servicios en línea. Este aparato, se sustenta, entonces, a través de la creación y gestión del registro único de la población, como es el caso de la “Clave Unica” en Chile, o el proyecto “CPF para todos”, en Brasil.

La gestión de los recursos escasos se vincula cada vez más a la ciencia, ya que al parecer tenemos una mayor afinidad con la producción de información relevante, pero es el Estado que siempre ha estado jugando este juego. No se puede negar que el Estado es la parte más interesada en la gestión de los recursos estables, la gestión del dinero público. y la redistribución de todos los beneficios, o sea, la ‘buena’ inversión de sus recursos. Y para maximizar esto, para ampliar su capacidad de gestión y control, los Estados se están dando cuenta de que necesitan ‘saberlo todo’. ¿Y cómo saberlo todo en un país tan grande como el nuestro? Patronizando, sandarizando a las personas. Metiendo a todo el mundo en su base de datos a través de un código identificados: su CPF. ¡Hasta acá nada de nuevo! Así funciona Hacienda, ¿no? Sólo que ahora, con internet, lo digital y demás, tienes la capacidad de

etiquetar y seguir a este tipo más fácilmente. No necesitas contratar a un investigador del IBGE para hacerlo, ir puerta por puerta, esas cosas. Ahora obligas al ciudadano a tener un CPF que opera dentro de alguna plataforma o servicio digital del gobierno. Y ya está. Así que ahora se están dando cuenta de que es mejor invertir en garantizar el acceso de la población a internet, en alfabetizarlos digitalmente, obvio! Conseguir que un niño de Paraíba matricule a su abuela en el programa Luz para Todos. A través de internet es más barato que enviar allí un equipo de asistentes del gobierno allá, ¿no?. (VPL - Entrevista 43/2023 - Brasil)

En la entrevista se destaca el modo en que el Estado ha desempeñado un papel clave en impulsar la digitalización de la sociedad, lo que contribuye directamente al movimiento de expansión de las implicaciones que conllevan el tecno-encanto. Se señala que para maximizar la capacidad de gestión y control estatal, una de las estrategias es que, precisamente, apoyen la inversión en el acceso a Internet y la alfabetización digital. Entretanto, la digitalización del Estado, no solamente se ha justificado por la modernización del estado y la optimización de los procesos burocráticos que involucran diversos sectores en pro de su funcionamiento más eficiente, sino que también es una estrategia de fortalecimiento y ampliación del desarrollo económico del país.

Al comienzo del texto se habló sobre la tipificación de la vida y la importancia económica de los datos que ha traído la contemporaneidad en la era de la información. Bien podría suceder que los datos tengan más relevancia que el petróleo, y ser lo que movilizara las industrias poderosas en el mundo. Teniendo en consideración esta realidad es que la conectividad nacional y la digitalización de la sociedad, es parte de la estrategia de inversión de las grandes industrias, con miras al crecimiento de la economía nacional, del presente, y sobre todo del futuro. Tener en manos los datos de la población entera del país significa una mayor capacidad de prever, predecir y adelantarse a asuntos como manejo de recursos energéticos, control de crecimiento de la población, inversiones de largo plazo, entre muchos otros. Cuanto mayor la capacidad de conocer, mayor es la potencialidad de mejorar la eficiencia social, e incluso mayor es la posibilidad de controlar. Controlar las personas, los recursos y lo que se espera del futuro, incluso para llegar preparado, más consciente de un panorama próximo.

Otro sector que desempeña un papel clave en la expansión de la conectividad global y la diseminación del tecno-encanto son las grandes empresas de tecnología y las compañías de telecomunicaciones. A menudo estas empresas invierten en infraestructura, como torres de

telefonía móvil y cables submarinos, para extender la conectividad a áreas rurales y remotas, en intercambio de instalar bases o explorar los recursos, inclusive de datos, de las regiones acordadas. Lo que aparentemente beneficia amplios sectores, ya que la conectividad global es fundamental para el crecimiento de la economía no sólo del país, como la microeconomía individual, a nivel de cada ciudadano, ya que la economía digital permite a las empresas llegar a nuevos mercados, brindar servicios en línea y aumentar su alcance, impulsando el crecimiento económico y la creación de empleo (Almeida, Dias, y Santos, 2021).

Es importante mencionar que a pesar de los esfuerzos, todavía existen desafíos financieros y logísticos para conectar a las poblaciones excluidas de la vida digital. Uno de ellos es que la expansión de la infraestructura de banda ancha puede ser costosa, especialmente en áreas remotas con una baja densidad de población. El segundo tiene que ver con la regulación y políticas públicas. Las políticas gubernamentales desempeñan un papel importante en la promoción de la competencia entre los distribuidores de internet, barateando el costo de su acceso a la población y son capaces de eliminar barreras burocráticas en este proceso, facilitando así la expansión de la conectividad, y con ella, la implantación del tecno encanto se ve fortalecida.

Como se puede ver, al fin y al cabo el imperativo de la conectividad global es un objetivo importante para los gobiernos y las grandes industrias debido a su impacto en la inclusión social, el desarrollo económico y la calidad de vida. Todavía, sin embargo, queda trabajo por hacer para reducir la brecha digital y garantizar que un mayor porcentaje de la población mundial tenga acceso a Internet y sus beneficios. Es una derivación directa de la percepción social y mundial con respecto a los beneficios que significa la digitalidad, las tecnologías computacionales, post internet, o sea, esto podría ser la consolidación del tecno-encanto.

Así como la incorporación de los procesos computacionales y digitales a la hora de investigar, genera en el ambiente de las ciencias sociales la expectativa positiva, es nítida la percepción que entre la comunidad de investigadores en las ciencia de datos hay un deslumbramiento aún mayor con relación a las potencialidades y posibilidades que la aplicación de los saberes en ciencia de datos proporciona y agrega a la investigación social aplicada.

Esta fascinación por la resolución de problemas de manera computacional o automática, está muy asociada a las promesas de innovación, mejoramiento y optimización de los procesos que la implementación que nuevas tecnologías suelen despertar en sus usuarios humanos.

También tiene orígenes en el desarrollo de las capacidades computacionales que hemos vislumbrado en la última década, del abaratamiento de los costos de adquisición de máquinas y equipos y de una cierta expansión de las capacidades de los distintos profesionales en introducir en sus actividades laborales, prácticas computacionales de procesamiento de información proveniente a que son expertos.

El aumento del uso de Tecnologías de la Información y Comunicación (TIC) nos brinda una sensación de privilegio y poder en comparación con otros seres en el mundo. Sentimos que podemos influir en nuestro entorno mediante la comprensión, manipulación y transformación que las tecnologías nos permiten. Adoptamos una posición de privilegio tanto con respecto a la naturaleza, como a nuestra relación con el mundo natural, del que creemos no ser parte.

El tecno-encanto también nos infunde la sensación de ser superiores debido al privilegio de adquirir y generar conocimiento especial sobre el mundo, incluso sobre lo que está por venir. Este fenómeno se debe tanto a la capacidad de comprender, descubrir y actuar de manera más rápida y efectiva gracias a la vasta base de datos, como también a que está construido a partir del proceso de automatización de la vida y de la facilitación de los procesos y actividades. Ello contribuye a la consolidación de la noción de que el futuro promete avances significativos, conquistas sobresalientes y descubrimientos novedosos. Tal escenario influye en la forma en que los seres humanos se relacionan con la tecnología, con la naturaleza, y con la sociedad.

De acuerdo a Latour (2008), el ser humano da forma al mundo a medida que es influenciado por lo que crea. Las grandes revoluciones de la humanidad, como el descubrimiento del fuego, la revolución agrícola, el darwinismo y la revolución científica, además de la navegación, la Revolución Industrial, la ley de la relatividad de Einstein, son apenas unos pocos ejemplos que nos muestran la forma en que una sociedad se construye, mantiene o se transforma en respuesta a la creación de dispositivos, y artefactos tecnológicos que se han concebido para comprender el mundo.

En la época actual no es diferente. La introducción de nuevas técnicas y tecnologías de comunicación y computación ha alterado no sólo la forma en que se hace ciencia, sino también la manera en que la sociedad se concibe a sí misma y la forma en que interactúa en cuanto a colectividad. La difusión masiva del tecno-encanto en una sociedad occidental, digital y conectada hacen que experimentemos un imaginario colectivo de expansión de las

capacidades del hombre sobre el mundo natural, establecida básicamente a partir de las creencias anteriormente mencionadas: la confianza continua en la evolución tecnológica; la creencia de mejoría constante del avance científico y de los propios artefactos tecnológicos, instaurando en general, aunque con cierto recelo, la creencia optimista en un futuro mejor construido en conjunto con los diversos aparatos tecnológicos, de inteligencia computacional que manejamos.

**Tabla 34.** *Resumen esquemático tecno-encanto*

Tecno-Encanto	Palabras Asociadas	Patrones Instaurados
	<ul style="list-style-type: none"> <li>* Encantamento</li> <li>* Fascinación</li> <li>* Desarrollo Constante</li> </ul>	<ul style="list-style-type: none"> <li>- Confianza en el tecnológico</li> <li>- Evolución Tecnológica Exponencial</li> <li>- Futuro Optimista: “o melhor estar por vir”</li> <li>- Avance Científico</li> <li>- Lugar de Privilegio del Humano</li> <li>- El mundo está mejor que antes</li> <li>- Tecnología resuelve los problemas</li> </ul>

### ***9.1.3 Automatización de la vida: eficiencia y aceleración de los procesos.***

Refiere a la predisposición hacia la implementación de sistemas automáticos en la realización de demandas de la vida cotidiana. Significa el proceso de sustitución de los seres humanos en las funciones mecánicas, repetitivas y procedimentales, reemplazados por servicios y automatizaciones que realizan el mismo tipo de trabajo, y muchas veces con un grado aún mayor de eficiencia.

La eficiencia, es el beneficio que más frecuentemente está asociado al escenario de la vida automatizada de la sociedad contemporánea. El supuesto aumento de la eficiencia del trabajo y de las actividades cotidianas, por su parte, contribuiría al incremento de la productividad,

ya que se supone que al liberar a las personas de los trabajos procedimentales, estaría creando nuevos espacios de trabajos con demandas más estratégicas, analíticas y/o creativas.

La automatización de la vida genera una priorización a la eficiencia también por las vías económicas: la tecnología ha permitido la automatización de tareas y procesos en muchas industrias, lo que a menudo reduce la necesidad de mano de obra costosa, lo que disminuye los costos operativos, abaratando el costo de productos de industrias globales.

Agilidad, eficiencia y ordenamiento de tareas son capacidades ligadas y esperadas de la sociedad contemporánea (computacional), y están contenidas en el imaginario de significados que contiene la conceptualización de techno encanto. Sumado a esto, la masificación de la utilización de programas de Inteligencia Artificial, y los sistemas de IA Generativa, como el Chat GPT, recién lanzado en marzo de este año de 2023, hacen que la automatización de las tecnologías y artefactos generen verdadera fascinación en sus usuarios, dentro y fuera de la academia.

A causa de la fascinación que la automatización ha despertado en la sociedad, hay una sobrevaloración de términos como “productividad”, “eficiencia”, “ganancia de tiempo”, “optimización de los recursos”, “padronización de los procedimientos”, “replicabilidad de los modelos”. Lo vemos más allá del mundo del trabajo, la economía, o instituciones. La vida cotidiana y doméstica también viene siendo inundada de procesos y vocabularios automatizados. Los alarmas de rutina, que coordinadamente anuncian que es la hora de tomar agua, o pararse de la silla para mover un poco las piernas. La aplicación que envía un mensaje estandarizado de feliz cumpleaños a tu compañera del trabajo, al lado de la otra aplicación que mantiene en día las cuentas de tu casa, porque programaste un envío automático de pago a cada una de ellas. Todo esto mientras suena la olla robot diciéndote que la sopa que habías programado hace una hora atrás, como por arte de magia, ¡está lista!

En el mundo científico, pasa algo similar: la eficiencia del proceso de automatización es percibido como una ventaja, pues facilitan la vida del investigador y adelanta, optimiza y reduce el tiempo invertido en actividades que “nadie hace tan bien como la máquina”, citando a una académica colombiana entrevistada. En tal caso, la científica se refería al proceso de investigación que llevaba sobre la percepción colectiva de la valorización patrimonial vigente en el centro de Bogotá (Urbina, 2021). Según ella, ningún humano tenía la capacidad de categorizar tan variados tipos de imágenes en tan poco tiempo. También hablaba de la

capacidad del sistema el cual, segundos después de subir las imágenes digitalizadas, categoriza, en cuestión de segundos, el tipo y año de la imagen, además de algunos previamente programados elementos escenográficos, para luego almacenarlos en su carpeta correspondiente.

Por otro lado, sobretodo en el mundo de la gestión pública, hay una importante controversia en cuanto a las tomas de decisión a partir de sistemas automáticos: si, por un lado, las automatizaciones son bienvenidas, algunos proyectos de automatización no suelen tener la misma buena recepción en todas las instituciones y servicios públicos. Por lo menos al principio. Hay una preocupación y cautela en cuanto al poder de la máquina en decidir, ayudar a decidir o influir en el proceso decisorio, sea de leyes, de acuerdos judiciales, sobre la designación de beneficios públicos en salud, economía, vivienda, educación, etc.

Cierta resistencia hacia la implementación no consciente de los procesos que implican el funcionamiento de la máquina pública, es deseada e inclusive muy recomendada en los diferentes servicios públicos con los cuales tuve la chance de interactuar a lo largo de estos 5 años de investigación. Al final, “estamos designando a los desarrolladores del sistema el poder de decidir a qué familia vamos a entregar una casa valorada en 100 millones de pesos”, decía un participante de un grupo focal realizado con diversos gestores de servicios públicos chilenos.

Frases como la de esta representante del ministerio de vivienda de Chile, con la cual pude interactuar gracias al ejercicio de mis funciones laborales en el GobLab, reflejan un “buen” tipo de reflexión cuanto al poder de la utilización de los sistemas de modelamiento computacional, los algoritmos y sistemas de automatización. La mencionada funcionaria expresaba la preocupación general, de prácticamente los 12 asistentes a esta reunión: de que al ceder el poder de decisión a los sistemas (y sus desarrolladores), se estaba delegando la capacidad de determinar beneficios importantes a miles de familias en todo el país. Esta preocupación pone de manifiesto la necesidad de una reflexión crítica sobre el tema, además de la implementación de salvaguardias para garantizar que las decisiones tomadas por los sistemas automatizados sean justas y éticas, evitando la concentración de poder en manos de unos pocos y asegurando la transparencia y la rendición de cuentas en el funcionamiento de la administración pública.

De hecho, esta es precisamente una de las atribuciones del Repositorio de Algoritmos Públicos de Chile de UAI/GobLab, del cual soy parte como miembro del comité editorial. Mapear, acompañar y garantizar el buen uso de las tecnologías computacionales para incidir de forma responsable y ética en la gestión de los datos y soluciones públicas en Chile, para el bien de la ciudadanía.

En mi experiencia, la resistencia de parte de los gestores no es infrecuente. Frente a la implementación de ciertas soluciones automatizadas, podría ser por sentir su autoridad profesional amenazada ante actividades y las decisiones sugeridas por la máquina. Sobre todo si son discordantes con las suyas. Este recelo hace conectar con el origen muy próximo a la discusión que cotidianamente vemos en las calles, películas y cada día más en los noticieros de la tv: la dominación del mundo por las máquinas, o el clásico embate de la ciencia ficción: “hombre versus máquina”. Con todo, en términos generales, aunque haya cierta desconfianza y recelo de la introducción de esta tecnología, la automatización de los procesos de trabajo e investigación son bien recibidos por la comunidad beneficiada por estos sistemas.

Un tercera derivación del tecno-encanto presente en la sociedad computacional, o un efecto colateral, es la promoción de la cultura del aprendizaje continuo, donde las personas reconocen la importancia de adquirir habilidades digitales y mantenerse actualizadas en un entorno tecnológico en constante evolución, para mantenerse a la par del avance tecnológico y el trabajo que involucra este saber. Logrando, también, mayor resiliencia a los cambios y rapidez en la adopción tecnológica. En otras palabras, una sociedad igualmente predispuesta a las innovaciones. O como diría un amigo mío, “vivimos en la era de la espera por la actualización aún más impresionante”. Esto se replica en la esfera del trabajo, con la necesidad de constantes y cortas especializaciones, cursos de operación de diversos softwares, participación en eventos y ferias en línea, en diversas otras áreas de la vida social.

La innovación y el cambio son una constante en la sociedad contemporánea, y la población se ha acostumbrado cada vez más a los cambios en sus dispositivos y las actualizaciones constantes de las numerosas aplicaciones necesarias para el funcionamiento cotidiano. Esta adaptación a los cambios continuos ha llevado a una mayor obsolescencia de los productos adquiridos, ya que rápidamente son superados por las nuevas y sorprendentes actualizaciones que se vuelven indispensables. Al mismo tiempo, la vida se ha vuelto más acelerada, y las personas se han acostumbrado a la idea de cambios constantes y frecuentes en su entorno.

Con respecto, por ejemplo, a proyectos en datos en el tercer sector, los bancos, supermercados y empresas de telefonía móvil, la automatización del proceso de análisis de sus datos está completamente incorporado en la cultura de las empresas. En una de las entrevistas realizadas con científicos sociales, encontré un magister en filosofía, de formación en comunicación social (pregrado), que llevaba más de 5 años trabajando como analista de datos para un banco multinacional con sede en Brasil. Según su relato, en su contexto laboral, las automatizaciones y el uso de programas automatizados era rutina, donde había más de 2 mil personas habilitadas y capacitadas para tal. A lo largo de la entrevista, menciona que el proceso de formación del área de datos del banco, era la segunda más grande en términos de inversión de recursos, demostrando, así, la importancia del continuo aprendizaje que implica la utilización de soluciones automatizadas, por parte de sus usuarios:

Prácticamente todo mi trabajo se hace automáticamente utilizando modelos de aprendizaje automático. En mi departamento no hay una sola persona que haga los cálculos. Hacemos predicciones basadas en la formación que han recibido los modelos. Publicamos estas previsiones en la plataforma Gutenberg. [...]. Ella es compuesta de varias partes, por ejemplo, una parte del desarrollo de modelo de aprendizaje automático en la que necesitamos un software que garantice el registro de los modelos, otra de seguimiento de los parámetros, de las métricas utilizadas en cada modelo, en cada sesión de entrenamiento... Y todo esto, estos tipos de herramientas, tenemos que ser capaces de proporcionar a los científicos de datos. Ofrecemos la parte de programación. Pero nos pasamos todo el día hablando de ciencia de datos, aprendizaje automático e inteligencia artificial, tenemos que saber de qué se trata, para poder hacer que esto funcione realmente dentro de nuestra plataforma y garantizar que los científicos de datos tengan esta comprensión, para que puedan trabajar con los datos. Todo esto dentro de una plataforma propia operacionalizada por más de 2 mil profesionales que tiene que capacitarse constantemente para tanto. Incluso para poder acompañar el ritmo de evolución del sistema. (PCL - Entrevista 09/2023 - Brasil - mi traducción)

La capacidad de adaptación a nuevas tecnologías y aplicaciones se refleja en la sociedad contemporánea, donde las personas adoptan con rapidez las últimas tendencias tecnológicas y se familiarizan fácilmente con nuevas herramientas. Como resultado, también se ha fortalecido su capacidad de resiliencia para afrontar desafíos tecnológicos, como la transición a nuevas plataformas, la solución de problemas técnicos y la adaptación a interrupciones

tecnológicas Esta habilidad para abrazar el cambio tecnológico ha sido una característica distintiva de la vida contemporánea.

**Tabla 35.** *Resumen esquemático automatización de la vida*

Automatización	Palabras Asociadas	Creencias Instauradas
	<ul style="list-style-type: none"> <li>* Eficiencia</li> <li>* Aceleración</li> <li>* Optimización</li> <li>* Replicabilidad</li> </ul>	<ul style="list-style-type: none"> <li>- Agilidad en los procesos</li> <li>- Padronización de los procedimientos incidiendo en menores tasas de errores</li> <li>- Resultados más rápidos y robustos</li> <li>- Minimización del trabajo brazal humano</li> <li>- Mayor capacidad de comparación</li> <li>- Aprendizaje continuo</li> <li>- Mayor ordenamiento y organización de la información y parámetros</li> </ul>

#### **9.1.4 Velocidad**

Como vimos en el apartado anterior, la automatización de los procesos de la vida, en la actual sociedad computacional, provoca, en cierta medida, el aceleramiento de los procesos en diversas esferas de la vida. Decisiones económicas que priorizan cortos plazos, investigaciones cada vez más acotadas en años de estudios, industria del consumo obligando a los consumidores a renovar sus equipamientos cada año, automatización de aspectos doméstico de la vida a través del aplicaciones de gestión del tiempo, ha contribuido de igual manera para la introducción del imperativo de la velocidad en la sociedad computacional.

El imperativo de la velocidad, de la rapidez, habla respecto de las consecuencias de una vida pautada por los parámetros de la comunicación en tiempo real y a la relativamente fácil accesibilidad a la información vía internet. Creamos a partir de estas dimensiones socio técnicas de la vida, un ambiente donde la instantaneidad pasa a ser considerada el patrón. La capacidad de la sociedad contemporánea para acceder, compartir y comunicar información de manera casi instantánea debido a las tecnologías digitales y las redes de comunicación generan una reducción de la expectativa del tiempo de respuesta a las diversas instancias de comunicación en las cuales convivimos, lo que a la vez, implica y exige mayor rapidez en la comunicación global en general.

Tal panorama también es responsable por la masificación y abaratamiento de los costos de internet y acceso a aplicaciones (aps) y programas simples de operar a un muy bajo costo. La información pasa, entonces, a ser distribuida de manera mucho más directa y autónoma entre las personas, con la acción de un *click*, sin grandes intermediarios, como era el caso con las tradicionales emisoras de televisión. Es justamente el quiebre de este intermediario y la propagación de información, lo que causa que la velocidad de la circulación de noticias, informaciones e incluso mundiales, se torne una barrera mucho menor comparada a, incluso, siete años atrás (Lemeshchenko et al., 2022).

La información en línea está al alcance de la mano a través de motores de búsqueda y aplicaciones, lo que permite obtener respuestas instantáneas a preguntas sobre el más variado asunto, y la capacidad de acceder a noticias, datos y recursos en cuestión de segundos, retroalimenta un proceso de identidad de la rapidez en la sociedad computacional contemporánea.

Compartir y difundir información rápidamente, comunicados, llamadas internacionales, fotos, audios, videos, memes, ubicaciones, contratos, a través de las plataformas de mensajería y redes sociales, permite a las personas compartir pensamientos, estilos de vidas y nuevas necesidades de forma instantánea, lo que puede llevar a la viralización de contenido en poco tiempo, a la vez que hay una constante retroalimentación de la información, en tiempo real y a escala global.

Otro sector de la sociedad que se ve favorecido por la rapidez e instantaneidad del proceso de obtención de información, es el de las empresas y organizaciones. Estos, pueden recibir retroalimentación de sus clientes o audiencia de manera casi inmediata a través de encuestas en línea, reseñas y redes sociales, lo que les permite tomar decisiones y ajustes rápidos, al tiempo que impacta directamente la disponibilidad instantánea de datos y su análisis en tiempo real. Es decir, pueden tomar decisiones basadas en información constantemente actualizada y precisa.

Como consecuencia de esta realidad, se percibe un proceso de retroalimentación del imperativo de la velocidad, ya que la sociedad de forma general, espera, siendo esta consumidores, empresas o funcionarios públicos, respuestas rápidas a sus mensajes, y consultas, procedimientos, retorno bajo los investimentos, cantidad de *likes* en las redes sociales, etc, exacerbando y difundiendo una cultura (de culto) de la inmediatez.

Incluso en contextos de investigación científica social, el principio de la velocidad es puesta en debate para respaldar, por ejemplo, la necesidad de automatizar o cambiar la manera en que tradicionalmente se ha obtenido información social para la realización de pesquisas. Uno de los entrevistados, al ser cuestionado sobre la validez de los métodos estadísticos frente a las investigaciones que están implementando a partir de técnicas en ciencia de datos, deja claro que la velocidad es un ítem a ser considerado en los contextos de investigación científica social:

El primer punto es tratar de entender que la investigación trata de representar el mundo, y el mundo está en constante cambio. Tal vez lo que está sucediendo no es que sean instrumentos anticuados u obsoletos [refiriéndose a tradicionales técnicas de muestreo de la estadística], pero la verdad es que las cosas están cambiando más rápido. Quizá tengamos que aumentar la velocidad a la que producimos interpretaciones de la realidad. Y esto no es un problema nuevo. Ya lo decía Adorno. Las ciencias sociales siempre están hablando de un mundo que ya no existe. Él lo decía en los años 1930. Los resultados electorales son fáciles de detectar, el 'error' porque las elecciones lo demuestran. Pero qué pasa con todas las investigaciones que hacemos, en las que no sabemos si acertamos o nos equivocamos porque el mundo no se revela y cambia cada 30 segundos!?! [risa] Ahora nos hace estar aun más retrasados. (WLM - Entrevista 12/2023 - Brasil. mi traducción)

El entrevistado insiste en la necesidad de comprender que la investigación pretende representar un mundo en constante cambio, que se transforma, por lo tanto, rápidamente. Sugiere que la aparente obsolescencia de las herramientas de investigación puede reflejar en realidad la velocidad del cambio de la sociedad (computacional) contemporánea. El entrevistado subraya la importancia de aumentar la capacidad de producir interpretaciones de la realidad con mayor rapidez, reconociendo que las ciencias sociales a menudo abordan un mundo que ya no existe. Es por esto que también menciona el reto de evaluar las investigaciones sociales cómo se está haciendo con los estudios de opinión electoral, donde los resultados ofrecen una clara verificación, en contraste con otras investigaciones cuyos resultados no son inmediatamente evidentes debido a la complejidad del mundo actual. Podemos ver, por lo tanto, en esta sociedad veloz, el deseo latente por acelerar el proceso o la capacidad que tenemos no solamente de producir información, sino sobretodo de analizarla, saber interpretar la diversa y gran cantidad de información cambiante, conectándolo con acontecimientos muy actuales, y muchas veces, la necesidad de ser capaz de hacer análisis de coyuntura en tiempo real.

En resumen, la instantaneidad de la vida es una característica de la sociedad computacional contemporánea, la cual experimenta el tecno-encanto, lo que ha sido habilitado en gran medida por la tecnología digital. Facilita la comunicación, el acceso a la información sin intermediarios, facilita la toma de decisiones en tiempo real, muchas veces incluso con alcance global.

**Tabla 36.** *Resumen esquemático de velocidad*

Velocidad	Palabras Asociadas	Patrones Instaurados
	<ul style="list-style-type: none"> <li>* Aceleración</li> <li>* Instantaneidad</li> <li>* Corto Plazo</li> </ul>	<ul style="list-style-type: none"> <li>- Mayor Facilidad de Comunicación</li> <li>- Percepción de aceleración del tiempo de respuesta</li> <li>- Deseo en acompañamiento en tiempo real</li> <li>- Necesidad de retroalimentación de la información</li> <li>- Constante actualización de la vida y capacidades</li> </ul>

### **9.1.5 Identidad global**

La secuencia anterior nos evidencia como la rapidez y el imperativo de velocidad, a partir de la instantaneidad de la vida, logra, como efecto colateral, eliminar los intermediarios de la comunicación, haciendo que la distribución de información alrededor del mundo sea más fluida y amplia, incluso globalmente. El quinto principio que se observa en la sociedad computacional es la identidad global, que emerge justamente de este proceso de interconexión global que junto al acceso a recursos en línea, facilitan la adaptación a las tendencias tecnológicas globales y la colaboración en proyectos a escala internacional. Esta colaboración está diseminada en diversas esferas de la vida social.

En la esfera económica acompañamos el crecimiento de la economía global e interconectada debido a que las pequeñas y medianas empresas, emprendedores e investigadores financieros están disfrutando de un escenario más abierto a inversiones en capital extranjeros, gozando de mayores oportunidades para llegar a mercados internacionales. Esto amplía el alcance y las posibilidades de crecimiento económico global en general. Otro importante impacto en la economía es el relativamente fácil acceso a los recursos globales. La tecnología ha permitido acceder de manera más rápida y sencilla a estos, ya que hemos experimentado un desarrollo muy impresionante de las tecnologías de entrega de producto, rastreo y seguimiento de mercancías. Las compras por internet a mercados internacionales como los de Estados

Unidos y China no paran de crecer, son unos de los que más lucran en el mundo post pandemia. El acceso a bienes y la contratación de servicios en varias partes del mundo, debido al comercio y mercado digital, nos otorga, en cuanto sociedad contemporánea, una vivencia de accesibilidad global, nunca antes vista.

La accesibilidad global en el contexto de la sociedad computacional también es responsable por el fortalecimiento del techno encanto y el deslumbramiento con el desarrollo tecnológico que experimentamos en cuanto sociedad global y interconectada.

Como ya he mencionado, los avances tecnológicos tiene importantes implicaciones en la vida cotidiana de la sociedad contemporánea, desde la manera cómo impactan las decisiones más domésticas, como la ruta elegida para ir al trabajo, a la mantención de la salud mental de una familia en situación de migración, simplemente por tener la posibilidad de hablar diariamente con sus familiares a un costo muy bajo.

La interconexión y el acceso a la información y recursos en todo el mundo también ha impactado la esfera de la salud, medicina y ciencia. La telemedicina, por ejemplo, ha proporcionado una satisfactoria atención a una parcela de la población que de otra manera no podría tener acceso a médicos o profesionales de calidad, sin salir de sus casas. También ha aumentado el número de colaboración entre hospitales e instituciones de salud, laboratorios de investigación científica alrededor del mundo, gracias a la posibilidad de reunirse en línea. Mi propia experiencia profesional evidencia esta característica: todos los proyectos de investigación en que he trabajado en este ciclo de investigación han sido compuestos por profesionales internacionales, viviendo en diversos países del mundo, originarios o migrantes. Incluso ya es posible realizar procedimientos quirúrgicos a distancia gracias a la conectividad y el desarrollo de las máquinas y tecnologías al servicio de la biomedicina.

La interconexión global ha disminuido las fronteras geográficas entre personas, un caso clásico de los análisis del globalismo de los años 2000, además de que ha permitido la colaboración formal e institucionalizada de los trabajos en línea, y el uso de plataformas colaborativas como Microsoft Teams, Discord, Trello, Google Space Work o Slack. Este recurso ha contribuido significativamente a la instauración de otro tipo de colaboración: los equipos trabajan juntos de manera autogestionada, disociada de la ubicación geográfica tanto de la institución, como de sus colaboradores.

El teletrabajo, la posibilidad de trabajar desde cualquier lugar, es una consecuencia directa de la interconectividad del mundo. Hay estudios que estiman que el teletrabajo no sólo ha ayudado a disminuir los costos de contratación de profesionales altamente cualificados, reducido los costos asociados con la infraestructura física de oficinas y viajes, lo que puede traducirse en ahorro de costos para las empresas y empleados y también de proporcionar la integración internacional de sus equipos; sino que también ha tenido significativos resultados en cuanto a la productividad al permitir que las personas gestionen su tiempo de manera más adecuada según sus necesidades, incluyendo, por ejemplo las biológicas, muy en función de la reducción de inconvenientes para desplazamientos, costo de vida y condiciones de trabajo.

También se percibe a través del contacto con diferentes e internacionales organismos de investigación, con los cuales tuve la oportunidad de interactuar, que hay una suerte de creatividad y más ánimo con relación a la solución de problemas, ya que la limitación geográfica ya no es una barrera, permite incluir a diferentes personas al ambiente de trabajo. Culturas, idiomas, países y formaciones diversas conviven, en el sentido de armar otra manera de colaborar y resolver los objetivos por los cuales trabajan.

La capacidad remota de trabajar y la identidad de colaboración en el trabajo, ha cambiado de manera significativa la forma como vivimos nuestra vida laboral (Weller, 2018). Aparte de la colaboración y comunicación internacional, las redes sociales de trabajo como LinkedIn, Academia.edu y ResearchGate, por citar algunos, además de las videollamadas y las aplicaciones de mensajería, permiten mantener una relación colaborativa con personas de todo el mundo, lo que enriquece la vida cotidiana laboral gracias a nuevos vínculos profesionales (Carrigan y Jordan, 2022; Van Noorden, 2014)

La internacionalización del trabajo también acaba por proporcionar nuevas instancias y oportunidades de aprendizaje, lo que se alinea recíprocamente al principio anterior del aprendizaje continuo. La accesibilidad global a recursos educativos en línea, como cursos y tutoriales, brinda oportunidades de aprendizaje en cualquier momento y desde cualquier lugar. Esto permite a las personas desarrollar sus habilidades y conocimientos de manera conveniente, diversa y participativa.

Hemos visto que la accesibilidad global significa el acceso a información, recursos y servicios de cualquier parte del mundo. Lo que, por un lado ha enriquecido la vida cotidiana al proporcionar acceso a una amplia gama de conocimientos, cultura, entretenimiento y

oportunidades laborales, pero que por otro lado, nos ha impregnado de un sentido de uniformidad de la vida.

Esta dimensión es responsable por diseminar una visión homogénea de futuro: los procesos automáticos y decisiones en base a sistemas computacionales provocan una suerte de estandarización de la vida, donde a pesar de la supuesta sensación de diversidad, además de la amplia posibilidad de expresión del yo individual, se observa menos espacio para la creación. Como nos explica una de las autoras entrevistadas:

Creo que la restricción del campo de las elecciones actúa en los dos campos: macro-político y micro-político. Estamos siendo guiados por algoritmos que predicen lo que vamos a comprar, el próximo clic, la próxima compra, lo que vamos a consumir como contenido en las redes sociales, y parece que estamos, con esto, perdiendo un poco la dimensión de creación de diferencia de la que hablan Deleuze y Guattari: cuanto más restringimos este campo de posibilidades, más dejamos que las elecciones sean hechas por máquinas, por algoritmos, y llegamos a un escenario, ya especulando más en el futuro, en donde se anhela la automatización total: dejar la elección en manos de las máquinas. Franco Berardi es otro autor importante, con un texto que se llama 'Después del futuro', que piensa también un poco en esta restricción del campo de posibilidades macrosociales, debido al mapeamiento de las sociedades, que afina estas elecciones, reduciendo las posibilidades del futuro hacia un futuro que ya está predeterminado por las máquinas, o sea, no es espontáneo y no es creativo. (MRT - Entrevista 34/2022 - Brasil - mi traducción).

El algoritmo, las automatizaciones y la masificación de los deseos y voluntades individuales y padronización o reducción del universo de elecciones (Toledo, 2020) determina y proyecta nuevos comportamientos que están orientados por elecciones anteriores o pre determinadas por la máquina, o por lo que sea que el algoritmo identifica como su preferencia. Se pierde una dimensión de la espontaneidad de la toma de decisiones (uno elige porque estuvo estimulado a elegir lo que el algoritmo determina lo que es más probable para ti), lo que no es menor en términos de discusión de los estudios de consumo, mercado y opinión pública, por ejemplo.

En resumen, independiente de la discusión sobre las implicaciones negativas o positivas de la implementación de una identidad global en la sociedad contemporánea, la accesibilidad global habilitada por la tecnología ha transformado la vida. Al ofrecer una amplia gama de

oportunidades, experiencias y conexiones que antes eran inaccesibles o limitadas por la geografía, el diario vivir se vio drásticamente modificado. Esta característica es fundamental en la sociedad contemporánea y contribuye a un mundo más interconectado y enriquecedor en términos culturales, económicos y sociales.

**Tabla 37.** *Resumen esquemático identidad global*

Identidad Global	Palabras Asociadas	Patrones Instaurados
	<ul style="list-style-type: none"> <li>* Globalización</li> <li>* Conexión</li> <li>* Tendencias internacionales</li> </ul>	<ul style="list-style-type: none"> <li>- Interconexión global</li> <li>- Uniformidad de la vida: masificación de los deseos</li> <li>- Cuantificación de los parámetros de vida</li> <li>- Internacionalización del trabajo</li> <li>- Disminución de las fronteras</li> <li>- Achicamiento del mundo</li> <li>- Oportunidades ilimitadas</li> <li>- Eliminación de intermediarios</li> </ul>

### **9.1.6 Alienación del trabajo humano por detrás de las máquinas**

Todas las dimensiones levantadas en el transcurso de este capítulo han discutido la manera en que el tecno-encanto ejerce una creciente fascinación sobre los miembros de la sociedad en la cual vivimos. Esta atracción se manifiesta a través de un marcado uso y dependencia de dispositivos tecnológicos, como smartphones, tabletas y computadoras, y su influencia se ve en múltiples aspectos de la vida cotidiana, así como es responsable por determinar la manera cómo nos constituimos como sociedad (computacional). Ahora bien, por detrás de estos artefactos y dispositivos varios, que responden a los principios de automatización, velocidad e identidad global, está la alineación del trabajo humano (Marx, 1884).

En su concepción más general, la alineación se refiere a la sensación de desapego, aislamiento o separación que una persona puede experimentar con respecto a sí misma, a la sociedad, a su trabajo o a algún otro aspecto de su vida. Sociológicamente, ha sido en gran medida influido por la obra de Karl Marx (1844) y otros teóricos de la sociología (Healy, 2020; Fuchs, 2014), y es desde acá que tomo prestado el concepto, para expresar otro rasgo fuertemente presente en la sociedad computacional bajo los parámetros que vengo planteando. Entenderemos la alineación como una intensificación de la falta de conexión, de entendimiento e incluso de reconocimiento de la existencia del trabajo humano por detrás de

las tecnologías y las soluciones producidas por las herramientas de procesamiento computacional de datos.

Los principios de automatización y eficiencia, velocidad e identidad global, hacen desaparecer, o que se minimice todo el trabajo humano involucrado en la operacionalización y funcionamiento por detrás de estas tecnologías. La sociedad computacional, entiende la operación de estas capacidades más bien como cosas de la maquinaria, o atribuyen gran parte de este esfuerzo a atributos únicamente tecnológicos, lo cual, en realidad, sabemos que no es todo lo que pasa.

Karl Marx (1844) definió la alienación en el trabajo como un proceso en el que los trabajadores se sienten desposeídos de la propiedad y del control de los medios de producción y, en consecuencia, se sienten alienados de su propio trabajo y de los productos que crean. En el contexto del capitalismo, los trabajadores a menudo no tienen control sobre el proceso de producción y no poseen los productos que producen, ya que estos pertenecen a los propietarios de los medios de producción. Esto lleva a una sensación de falta de realización personal y a una desconexión con el trabajo y el producto final.

Es importante destacar que aunque la teoría de la alienación de Marx se centra en el contexto del trabajo en una sociedad capitalista del 1800, la apropiación de este término todavía es apropiada para describir estos sentimientos de desconexión y aislamiento que el uso de las tecnologías computacionales despiertan en los usuarios, aun cuando estemos tan apegados a ellas. El concepto de alienación ampliamente trabajado por Marx, podría ser sacado del contexto industrial y aplicado al contexto de la sociedad computacional presentado directa o indirectamente a través de los trabajos investigativos en algunos de los textos (Toledo, 2020; Téllez Carvajal, 2020; Sabino et al., 2020.) o en las conversaciones provenientes de la entrevistas con los autores de estas publicaciones.

La obsesión por la innovación y por los procesos computacionales, acaban haciendo desaparecer el trabajo humano por detrás de cada modelo, algoritmo construido y decisión ejecutada. Los propios artefactos, las máquinas y el procesamiento automático y computacional acaban llevándose el crédito por el trabajo realizado, aumentando aún más su evidencia y relevancia en la vida social. Pero no se puede olvidar que sin el trabajo involucrado de todo el conjunto de personas trabajando en distintos frentes de la investigación, es imposible de ser realizado.

El distanciamiento de la toma de decisiones y de la implementación de decisiones creativas es una de las consecuencias de la masificación del principio de alienación del trabajo humano, acá presentado: en una sociedad altamente tecnologizada, donde los algoritmos y las máquinas desempeñan un papel cada vez más importante en la toma de decisiones, existe un riesgo de que los trabajadores humanos sientan que sus habilidades creativas y su juicio sean menos valorados. La implicación directa de este tipo de fenómeno es el recelo humano de volverse obsoletos y ser reemplazados por tecnología más eficientes que la capacidad de pensamiento y decisión humana.

Un segundo y muy marcado tipo de desplazamiento proveniente de tal proceso de alienación es la despersonalización de las relaciones laborales: la comunicación digital y la automatización de tareas pueden llevar a una despersonalización de estas. Las interacciones cara a cara y la colaboración humana pueden disminuir, lo que a su vez puede afectar la calidad de las relaciones en el entorno de trabajo, tal como se pierde un cierto grado de espontaneidad con respecto a las soluciones, ideas y arreglos, creados en ambientes donde las relaciones humanas son más presentes.

En un entorno tecnológico complejo y altamente especializado, como es el caso de las actividades basadas en el procesamiento computacional de datos, es posible percibir otra dimensión de la alienación de la participación humana en estos procesos: la externalización de las responsabilidades, es decir, que los actores humanos trabajadores se sientan apartados de las consecuencias reales de su trabajo y de su relación con la totalidad de los servicios prestados, o actividades desarrolladas. Por ejemplo, los programadores de software pueden no ser plenamente conscientes de cómo su código afecta a las personas o a la sociedad en general. Así como las personas pueden no entender cómo el simple hecho de vivir sus vidas digitalmente pueden determinar su propia forma de vivir y tomar decisiones en el futuro. Como consecuencia, estas situaciones acaban por generar una división del trabajo y una especialización extrema donde las tareas se fragmentan en partes cada vez más pequeñas y especializadas, limitando, así como se ha dicho, el entendimiento del todo en su desarrollo profesional. La cita abajo demuestra el alto grado de especialización profesional de las contrapartes técnicas de los proyectos sociales que involucran al entrevistado, científico social autor de una de las publicaciones en Scielo:

Hasta hace unos 10 años sólo había científicos de datos. No existían los ingenieros de datos ni los ingenieros de aprendizaje automático. Entonces, el científico de datos tenía que hacer todo: descargar los datos, limpiarlos, analizarlos, crear los modelos de aprendizaje automático y producirlos. Así que tenía que hacer todo este ciclo por su cuenta. Pero con el tiempo esto ha cambiado y, en gran parte debido a la influencia de las empresas, estas funciones se han especializado en torno a cada una de estas etapas. La división de tareas de los científicos de datos acabó dividiéndose en diferentes ocupaciones, relacionadas a las fases del trabajo del área de datos. Ya, entonces esa parte inicial de recopilación de datos, limpieza de datos, creación de algunos *pipelines*, ya nadie tiene que hacerlo manualmente: con un *script* puedes hacerlo todo tú mismo. Pero para procesar el *script* igual se necesita a alguien ahí, ¿cierto? Este es ahora el trabajo del ingeniero de datos. Trabaja mucho con código, por eso se le llama ingeniero de datos. Porque tiene muchos pies en el área de datos y en el área de programación: creará mucho código para automatizar toda esta recopilación y limpieza de datos. ¿Por qué? Para que cuando llegue al científico de datos, el científico de verdad, sólo se ocupe de analizar, crear modelos, entrenar modelos, hacer experimentos, pruebas, ¿sabes? El científico de datos es literalmente un científico. El ingeniero de datos prepara los datos. El científico de datos está allí probando y pensando qué hacer con todo eso. Ahora está todo dividido. Es difícil encontrar a alguien que trabaje como en los viejos tiempos. Haciendo todo, desde el principio hasta el final del proyecto (PCL - Entrevista 09/2023 - Brasil - mi traducción)

Esta es apenas una de las muchas demostraciones que podría haber con respecto a las consecuencias del proceso de profundización del proceso de alienación laboral apuntado por Marx (1844) hace tanto tiempo atrás. Percibimos una alienación con respecto a los productos, es decir no hay ciencia o control de los productos y soluciones que generan, lo que lleva a una sensación de desconexión con el resultado de su trabajo. Esto es la alienación del proceso de trabajo: la sociedad no entiende de forma general lo que pasa por detrás, o como se organiza, o cual es efectivamente el impacto que el uso de estos sistemas engendra en sus vidas; hay además una alienación con respecto sí mismos, debido a la falta de control sobre la construcción de la propia identidad, sentido de realización, creatividad y espontaneidad, minimizados por el principio de automatización y eficiencia.

El desarrollo de la industrialización y del capitalismo siguen sosteniendo la tesis marxista de la alienación del trabajo humano que implica la introducción de tecnologías, independiente de cuáles sean: la máquina permite al capital reducir el trabajo humano y el gasto de energía al mínimo, pero sólo con el fin de realizar un máximo de trabajo. En este proceso perdemos la conexión con el trabajo que conlleva la construcción de los que se hace junto a la máquina (Marx, 1972). Esta alienación se repite y se profundiza en la sociedad computacional.

**Tabla 38.** *Resumen alienación del trabajo humano*

<b>Alienación del Trabajo Humano</b>	<b>Palabras Asociadas</b>	<b>Patrones Instaurados</b>
	* No reconocimiento * Desconexión	<ul style="list-style-type: none"> <li>- Dependencia de dispositivos Tecnológicos</li> <li>- No separación del trabajo humano y el trabajo de no-humanos</li> <li>- Minimización de procesos humanos por detrás de las máquinas</li> <li>- Desconexión entre el trabajo y el producto final</li> <li>- Desvalorización del trabajo humano</li> <li>- Despersonalización de las relaciones laborales</li> </ul>

### **9.1.7 Identidad Anfibia**

En el vasto territorio de los estudios constructivistas de la Ciencia, Tecnología y Sociedad (CTS), sobre todo los que proponen las disciplinas de sociología y antropología, se ha adoptado el término *hibridismo* para ilustrar las implicadas conexiones entre ciencia, tecnología y sociedad (Latour, 2007; Haraway, 2009; Callon et. al., 2009), Este concepto ha sido la brújula conceptual que ha guiado la exploración de cómo estas tres dimensiones se entrelazan, modelan sus trayectorias y mutuamente se influyen.

En este enfoque, el hibridismo se refiere a la mezcla y la interconexión de elementos de la ciencia y la tecnología con factores sociales, culturales y políticos. No se trata simplemente de la coexistencia de estas dimensiones, sino de su integración y coevolución. Desde la perspectiva constructivista, se entiende que la ciencia y la tecnología no son entidades aisladas, sino que están moldeadas por las percepciones, valores y decisiones de la sociedad en la que se desarrollan. El hibridismo, en este contexto, destaca cómo estas entidades híbridas emergen a través de la interacción continua y la co-construcción entre la ciencia, la tecnología y la sociedad (Latour, 2007). El enfoque constructivista en los estudios CTS reconoce la importancia de considerar el contexto social y cultural en la comprensión de la ciencia y la tecnología, y cómo estas influencias mutuas dan lugar a formas híbridas de conocimiento y práctica.

En otros campos del saber, por ejemplo en biología, en la ciencia de la computación y sistemas o en lingüística este término también es frecuentemente utilizado. En biología, un híbrido es el resultado del cruce entre dos especies diferentes, ya sea natural o mediante la

intervención humana (Gardner, Simmons y Snustad, 1998). En tecnología, se puede hablar de híbridos para referirse a dispositivos o sistemas que combinan características de dos o más tecnologías diferentes (Backels, 2015). En lingüística, el hibridismo ocurre cuando se mezclan elementos de dos o más idiomas, dialectos en una misma expresión (Josiah, 2014). En antropología tenemos el concepto de García Canclini (2006) de cultura híbrida exponiendo sobre las formas en que las culturas contemporáneas se han vuelto híbridas debido a la interacción constante entre lo local y lo global, lo tradicional y lo moderno, entre innumerables otras utilidades conceptuales alrededor de este mismo término. En resumen, el hibridismo implica la fusión o combinación de elementos diversos para crear algo único y, a menudo, más complejo o interesante.

Sin embargo, durante mi proceso de investigación y análisis de la información producida, al intentar vincular los hallazgos con el interior del laboratorio y actividades investigativas me deparé con una paradoja conceptual: aunque el hibridismo ofrezca un marco valioso para comprender las implicaciones sociales de mi estudio, y endosar el correlato encontrado en el laboratorio y la sociedad, descubrí, y además definiendo, que la metáfora de que experimentamos un “mundo híbrido, una identidad social híbrida”, planteada por varios autores constructivistas, no captura completamente la especificidad única de la experiencia que vengo tratando de demostrar a lo largo de este trabajo. La realidad que se despliega a partir de mis hallazgos no se acomoda correctamente a la “fusión homogénea” que se supone implicar el hibridismo, tal como venimos estudiando.

Más bien, la realidad que vi revela una adaptabilidad y una capacidad de transitar entre distintos mundos, lo que nos recuerda más a la vida de un ser anfibio.<sup>113</sup> Este concepto que planteo, *identidad anfibia*, se presenta como una metáfora, en un intento de resaltar la flexibilidad y la adaptabilidad que hoy son necesarias para moverse fluidamente entre mundos hasta hace poco considerados muy distantes e incompatibles.

---

<sup>113</sup>La idea de este concepto nació en una entrevista con uno de los autores de las publicaciones académicas de las bases de datos estudiadas. El análisis y relectura de esta entrevista fue la que me hizo ver las implicaciones conceptuales que podría tener la acción de este concepto, en detrimento del hibridismo, que venía trabajando hasta entonces. La presentación de este concepto, de identidad anfibia, también hace parte de un proceso de decisión en cuanto a la metodología sensible difundida por Knorr-Cetina (2005), quien dice que “es preciso dar la voz” (p. 36). Este concepto proviene, entonces, no sólo de escuchar a mis sujetos, también de la decisión de utilizar términos, expresiones y metáforas nativas tal como se presentaron. Dar relevancia a los vocabularios presentes en el cotidiano de la investigación es parte de la introducción de la metodología de investigación sugerida.

Como el agua y la tierra, el conocimiento social y computacional conviven de manera armónica en un sistema (sociotécnico) que muchas veces se comporta como un ser de características anfibias: un ser que transita entre dos mundos, siendo lo que es, sin necesariamente fundirse en una nueva especie. El contexto múltiple presente tanto en los artefactos, los múltiples actores involucrados y en las características del trabajo que se imprime en dicho ambiente, generan un organismo, un sistema sociotécnico único, sí, pero no fundido. Adaptado específicamente para navegar en los distintos y pre determinados contenidos/áreas de interés, en este caso, de las ingenierías computacionales y los estudios de la sociedad.

En este análisis, propongo que la sociedad computacional, encontrada a partir del análisis de la información producida, en lugar de seguir el principio de hibridismo, está gobernada por un principio de identidad anfibia. Este enfoque expresa y resalta la capacidad de la sociedad para transitar entre los dominios del “mundo natural/exactas/duro” y “mundo social/cultural/blando”, pero manteniendo su identidad única en cada uno de ellos, en lugar de fundirse en una amalgama indistinguible de ambos componentes. La identidad anfibia preserva la capacidad de diferenciarse, de tener los abordajes y contenidos separados y demarcados. Por otro lado, sin duda, habla sobre la capacidad de construir puentes de comunicación entre dos polos, generalmente opuestos.

Así, es que presento el último principio característico encontrado en la sociedad contemporánea: la metáfora del anfibio. Esta, encaja muy bien con la idea de transitar entre dos entornos distintos, el contexto social y el contexto computacional en este caso. Expresa la capacidad de los proyectos de investigación social a partir de la ciencia de datos en colaborar, relacionar y transitar por ambos mundos con la finalidad de lograr sus objetivos y aprovechar lo mejor de dichos saberes y aplicabilidades.

Al igual que los seres anfibios, estos proyectos poseen adaptaciones específicas para navegar tanto en el mundo del conocimiento social como en el computacional, muy similar a las capacidades que estamos adquiriendo en cuanto sociedad. La sociedad computacional es una sociedad que vive una vida anfibia, guiada por automatizaciones, principios de velocidad, identidad global, y alienación del trabajo humano, sumersa en los procesos autónomos y espontáneos de la vida real a ser vivida por un ser humano no completamente guiado por las máquinas.

Hay un movimiento de preservación de la identidad del alma, de lo que es humano, en una tentativa de diferenciarse de aquello que es máquina, artificial, conviviendo muy bien con la dependencia a estos mismo artefactos. El mundo post pandémico, en la misma medida que nos sorprende y nos causa fascinación, por el desarrollo tecnológico y practicidad, también nos trajo, no mucho más tarde, la necesidad del retorno a lo análogo, un contra-movimiento digital de la domesticidad de la vida (Ben-Zvi y Luftman, 2022): las ganas por estar presencial, por los encuentros fortuitos y por una vida fuera de las pantallas, por lo menos en la esfera más privada.

La sinergia entre estos dos enfoques crea algo único y específico en esta sociedad. Finalmente, esta capacidad de transitar entre dos contextos aparentemente antagónicos sin perderse, es lo que hace un ser anfibio único en su especie.

Aunque el término híbrido también sugiere la combinación de dos elementos distintos, el de anfibio resalta la capacidad única de transitar entre esos contextos de una manera fluida y adaptativa. La necesidad de adoptar este concepto fue aún más endosada mediante al proceso de observación y seguimiento de mis sujetos de investigación. La diferenciación entre híbrido y anfibio, fue siendo construida principalmente a partir de la identificación de algunas principales diferencias manifiestas durante el periodo etnográfico de este estudio.

La primera de ellas habla sobre las nociones de adaptabilidad y flexibilidad: mientras híbrido implica la fusión de dos elementos, anfibio destaca la capacidad de adaptarse y funcionar de manera efectiva en ambos entornos. Los seres anfibios no sólo combinan características acuáticas y terrestres, sino que también tienen la flexibilidad para moverse entre ellos según sea necesario. Esto refleja la naturaleza dinámica y adaptable de los proyectos de investigación que combina conocimientos sociales y computacionales. Los proyectos de investigación en ciencia de datos y temas sociales no son simplemente una mezcla homogénea de conocimiento, sino que tienen la capacidad de sumergirse en el conocimiento social o tecnológico según las demandas de la investigación, generando, como ya señalado, un ambiente interdisciplinario.

El segundo es sobre la manera en que se da la preservación de su singularidad: los seres anfibios son únicos en su capacidad para transitar entre dos mundos, y cada especie tiene sus propias adaptaciones específicas. Del mismo modo, cada proyecto de investigación puede considerarse único en su capacidad para combinar datos, metodologías, técnicas y

conocimientos de manera singular, con objetivos específicos, sin que necesariamente la producción de las soluciones, resultados o conocimientos finales sean transdisciplinarios o modificados al interior. Existe un intercambio de información, de modos de funcionar e incluso la absorción de algunos parámetros de análisis e indicadores de investigación, pero no se conforman, como resultado, en un proyecto nuevo, sino en un saber fundido en una nueva esencia de trabajo.

Considero el mundo anfibio como mejor metáfora visual y descriptiva para elucidar respecto a las características de la sociedad computacional, por evocar imágenes de transición suave y de movilidad entre dos mundos típicamente opuestos: vivir en una sociedad de identidad anfibia significa estar propenso a desarrollar la capacidad de hablar diferentes idiomas, sin perder el acento del pueblo natal. Es también estar inclinado a encontrar los caminos, las maneras de transitar con fluidez entre contextos muy diferentes y que se conectan de maneras no convencionales. Es estar impactado por las noticias de todo el mundo, de los acontecimientos de relevancia global, y segundos después, superar el malestar al mirar la sonrisa de un gatito en la parte más privada de la casa de un desconocido de internet.

Transitamos entre lo mundial y lo estrictamente particular con gran facilidad. Trabajamos en espacios cada vez más hegemónicos y globalizados, al tiempo que tratamos de no perder la unicidad de la empresa y el sentido de lo que nos diferencia del resto del mercado. Transitamos entre el hombre orientado por la máquina, y el hombre que de ella se desconecta para sentirse más vivo, en la tentativa de constantemente equilibrar el tiempo que pasamos en la tierra y en el agua. En la sociedad computacional no hay un mundo en oposición, tampoco un mundo enteramente compatibilizado, mas si un mundo en conversación bilingüe e intencionalmente determinado. No es muy difícil evidenciar tal escenario en diversos contextos, no solamente en la vida de un laboratorio de ciencia de datos, también en la vida social .

Uno de los fenómenos más reveladores que evidencian la identidad anfibia de la sociedad computacional se manifiesta de manera especialmente evidente en el ámbito laboral. En este universo, los *profesionales anfibios* emergen como figuras altamente valoradas, destacándose por su capacidad única para transitar entre una multiplicidad de polos de conocimiento.

La complejidad inherente a un trabajo anfibio se manifiesta a través de la propia identidad anfibia. Se observa la presencia de profesionales altamente cualificados, quienes, lejos de

confiarse únicamente a su campo de saber tradicionales, contribuyen activamente en otros contenidos, abordajes o actividades asociadas a sus trabajos. Son también profesionales capaces de relacionarse con diferentes tipos de instituciones. Están capacitados para el trabajo empresarial autónomo o corporativo y transitan por ambientes académicos, ya que son altamente especializados y también a veces, mantienen relaciones muy cercanas entre sí, ya que prestan servicios como consultor con organismos públicos y servicios gubernamentales, los que acaban incorporando a sus actividades mano de obra externa en la realización de sus proyectos al interior de la máquina estatal. Esta transición se materializa a través de conexiones y recursos que se entrelazan entre esferas gubernamentales, académicas y del tercer sector, desafiando las fronteras convencionales del empleo. Estas exigencias apuntan hacia un mercado de trabajo que prefiere profesionales cualificados, pero además, diversos: que realicen múltiples tareas y se sepan posicionar en diferentes tipos de ambientes profesionales. Se busca flexibilidad en cuanto a la integración de nuevos frentes de trabajo y que sean capaces de contribuir a los esfuerzos colectivos asociados a su actividad.

Este fenómeno desdibuja las divisiones tradicionales de la sociedad basadas en habilidades, en un claro contraste con el esquema fordista de operacionalizar el trabajo. En la sociedad computacional, la valoración recae no en la expertise o la titularidad de los profesionales, sino, más bien, en la capacidad de cooperar, de entender y ejecutar determinadas técnicas de trabajo y en la habilidad en ejercer múltiples tareas. Es necesario decir que la colaboración, en este contexto, no implica la ausencia de autoridad o jerarquía de cargos, ya que, al estar vinculados a esquemas empresariales o servicios gubernamentales, la existencia de roles y posiciones sigue siendo una constante. Sin embargo, la naturaleza de la negociación y la forma de trabajo adopta una estructura más horizontal, con varios líderes operando con autonomía en sus respectivos sectores, fomentando una toma de decisiones distribuida y adaptativa. Este enfoque promueve una mayor flexibilidad y agilidad en la respuesta a las dinámicas demandas del entorno laboral contemporáneo.

La amplia diversificación de funciones, conocimientos, contextos y métodos de trabajo, insta a que los profesionales anfibios adopten enfoques multidisciplinarios y propongan soluciones más flexibles para abordar desafíos laborales. De hecho, tal flexibilidad puede manifestarse como una suerte de *bricolaje*, concepto mencionado por Knorr-Cetina (2005) al inicio del libro que guía esta investigación. En el contexto brasileño, el término *gambiarra* se escucha con frecuencia para describir las características de trabajos y soluciones derivadas de los

numerosos desafíos asociados con la conformación de un entorno laboral multidisciplinario, así como de las capacidades operativas e infraestructurales necesarias.

En Brasil, se relata que la implementación de tecnologías en macrodatos y la estructuración eficaz de un data center, por ejemplo, están estrechamente vinculadas a la capacidad de adquirir equipos, contar con personal capacitado para establecer conexiones entre máquinas y sistemas, realizar la instalación de programas y licencias, así como operar el equipo necesario para llevar a cabo todo este proceso. En un contexto de emergencia tecnológica, donde las soluciones están recién disponibles a bajo costo, se observa la consolidación de escenarios, laboratorios y estructuras de trabajo con datos que llevan a lo que los brasileños denominan *gambiarra*: "Tenemos que ingeniar, encontrar una solución, sortear el sistema para obtener una licencia más económica, pedir prestado a otro departamento para aprender lo que están haciendo allí. Todo centro de datos comienza como una gran gambiarra" (Comunicación personal, 2023). Demasiadas condiciones específicas en un contexto de emergencia tecnológica recién disponibilizada a bajo costo, implican la consolidación de escenarios, laboratorios y estructuras de trabajo computacional de datos, que conlleva a ambientes naturalmente más creativos.

Sin embargo, el resultado positivo de esta realidad se manifiesta en que, en este entorno, se permite que diversos profesionales converjan en torno a temas específicos, superando las barreras tradicionales de la especialización y promoviendo un intercambio dinámico de ideas y habilidades. La flexibilidad y apertura a las demás técnicas y conocimientos, es vista no apenas como una herramienta de trabajo, también como una gran oportunidad profesional para destacarse en el mercado:

Yo creo que lo que hace de mí un profesional raro es el hecho de que yo siempre he sido muy abierto a todo, yo he sido de las ciencias exactas, he sido muy matemático, me ha encantado ser docente y académico, pero también ponerlo en práctica, he trabajado en empresas como asesor. Digamos que, con mucho respeto, tengo una gran versatilidad, porque puedo vivir perfectamente desde la academia, ser docente, ya lo viste, y puedo ir a una empresa, asesorar y decirle 'no haga esto, hágalo así', y explicárselo como a un bebé, [*rien*]. O meterme a programar y decir 'yo lo hago'. Entonces, esa combinación es difícil de conseguir. En general los ingenieros somos muy psico-rígidos, autistas, medio introvertidos algunos, entonces no saben transmitir el mensaje, solamente saben echar el código y no saben explicar, no saben pensar más allá de lo que fueran entrenados, entonces, es difícil. Pero veo que cada vez más esta capacidad es más requerida. Si nosotros ingenieros duros no empezamos a desarrollar estas otras habilidades, vamos a ser destinados más y más a hacer trabajo muy pajero y

mecánico. Para destacar es necesario ser más maleable, ¡sobre todo si quieres ganar más plata! Y es por esto que fue tan difícil de encontrar, no es verdad. Aun estamos recién empezando con esta visión que te estoy comentando, mientras mis colegas duermen, ¡voy aprovechando! (RNN - Entrevista 41/2022 - Chile)

No obstante, no puedo afirmar que esta sea una realidad en contraste con el resto del panorama laboral mundial. En este estudio específico, no se profundizó en comparaciones entre el trabajo en laboratorios latinoamericanos y el de los países del Norte global. Es más, entre los científicos con los cuales me relacioné existía una tendencia a relatar que esta situación de bricolaje en el ámbito laboral computacional es resultado de una condición de dependencia que las regiones del sur global tenemos: menor capacidad tecnológica, profesionales menos cualificados y menores incentivos para incrementar la conectividad con la cual se requiere la mantención de las condiciones apropiadas. Esta aparente desventaja, sin embargo, nos fortalece en términos de innovación y creatividad. La forma única y llena de *gambiarra*s de abordar el trabajo con grandes datos en Latinoamérica impacta positivamente en la calidad de los resultados. A pesar de las dificultades en las llamadas conexiones transcientíficas (Knorr-Cetina, 2005) desfavorables, la manera peculiar y creativa en que abordamos los grandes datos en esta región contribuye a la generación de resultados innovadores. Esta peculiaridad refleja la capacidad de hacer frente a las dificultades uniendo diferentes conocimientos en pro de hacer que las cosas funcionen de manera eficiente y lógica, lo que por un lado valoriza y refuerza la característica de la transitoriedad y la habilidad de combinar elementos diversos para lograr un propósito común. Característica muy similar a otras investigaciones respecto de la producción de conocimiento científico latinoamericano (Feld y Kreimer, 2019).

Una característica sobresaliente que respalda la consolidación de una identidad anfibia en la sociedad computacional es la habilidad para adaptar las realidades locales a las crecientes demandas globalmente estandarizadas. La uniformización que la información adquiere mediante su empaquetamiento digital, permite que los análisis macrosociales de grandes y diversos datos puedan ser revelados mediante unos simples pasos o comandos de script: la especificidad de la información a nivel individual, desentrañando detalles hasta la línea específica en una tabla de Excel.

La maleabilidad de los datos digitales y sus variadas formas de procesamiento computacional posibilitan el tránsito fluido entre el ser generalista y realizar estudios macro a nivel global y comparativo, hasta la capacidad de examinar minuciosamente los datos, siendo muy específico en la obtención de información de un individuo en particular. La ciencia de datos permite hacer *zoom in* y *zoom out* con relativa facilidad en una base de datos, analizando variables y trazando relaciones tanto a nivel del total de la población como a nivel individual, grupal o de sistemas muy pequeños.

Esta realidad se confirma especialmente en la discusión sobre la infraestructura y operación de datos, donde se evidencia la necesidad de transitar entre una base global de datos, predominantemente dominada por el Norte global, y estructuras basadas en recursos y fuerza de trabajo locales. Aunque la mayoría de las bases de datos y del capital humano están situados localmente, se trabajan a partir de estructuras y recursos de operación globalmente estandarizadas y establecidas. Esto es ilustrado en una de las entrevista realizadas con uno de los colaboradores de una importante y pionera organización latinoamericana, Derechos Digitales.org:<sup>114</sup>

Entonces, en esta ocasión [las agencias auspiciadoras al percibir la expansión internacional que venía creciendo en la organización] nos entregaron dos opciones: desescalar el crecimiento de las organizaciones en cada localidad y que se volviese una cuestión más chica, o expandirse y se optó por expandirse. Y, con ese cambio importante dejamos de trabajar solamente en Chile, lo cual significó que dejamos de tener personal solamente en Chile. De la mano de eso, se contrató a la primera persona en México. Y de ahí ha sido como un periodo de expansión, que es territorial, tiene que ver con tener gente en distintas partes y estar atento a qué está ocurriendo en distintas partes. Un trabajo que es básicamente de generación de redes. Estamos trabajando acá, pero la idea siempre ha sido como generar posibilidades de colaboración con organizaciones locales que, probablemente, tienen menos experiencia, pero tienen conocimiento del territorio, tienen conocimiento de la realidad local, y, en el fondo, de esa unión, hacer sinergia para poder generar cambios más relevantes. Sin dejar también de prestar más atención a lo que está pasando en los organismos internacionales y las distintas discusiones que se están dando en el marco de los países más avanzados. (VGD - Entrevista Compartida - 23/2023 - Chile).

---

<sup>114</sup> <https://www.derechosdigitales.org/>

Como se puede observar, este comentario sugiere una dinámica compleja en las organizaciones de la sociedad computacional. La actuación a nivel local se ve impulsada por un deseo de fortalecer internamente al país o regiones. Sin embargo, existe una clara orientación hacia paradigmas internacionales, guiada por la necesidad de cumplir con normativas y acercarse a estándares internacionales. Este enfoque no sólo tiene implicaciones legales, sino también estratégicas, ya que busca captar atención, influencia y recursos provenientes del extranjero, que a menudo financian las actividades de estas instituciones.

El crecimiento y la expansión internacional son vistos como estrategias para obtener recursos y reconocimiento. En algunos casos, las agencias auspiciadoras ofrecen opciones que involucran decisiones clave, como desescalar el crecimiento o expandirse. La decisión de expandirse a menudo conlleva la contratación de personal en nuevas ubicaciones por la necesidad de tener representación cercana al norte global. Esto implica adaptarse a protocolos y requisitos específicos, además de aprovechar la proximidad para implementar actividades y mantener una conexión directa con las fuentes de financiamiento en Europa o Estados Unidos. Además del modelo y ejemplo, la presencia en el norte facilita el cumplimiento de protocolos comunes, esenciales para acceder a los fondos financieros que sustentan estas instituciones.

La expansión territorial implica estar atentos a las dinámicas de diferentes regiones y trabajar en la generación de redes. Esto significa no sólo realizar trabajos en el lugar de origen, sino también establecer colaboraciones con organizaciones locales. Este enfoque híbrido, combinando la atención a lo local y a la participación en discusiones internacionales, refleja la necesidad de equilibrar las realidades locales con los estándares globales. Además, demuestra la importancia de la colaboración internacional y de la generación de redes para las instituciones que buscan influir y generar impacto más allá de sus fronteras.

En términos específicamente de análisis del mundo científico y de la producción de ciencia es observable, en términos generales, que los proyectos de investigación, a pesar de seguir modelos internacionales, aplican sus conocimientos a la realidad local. Este movimiento constante entre influencias globales y adaptación a la realidad local refleja, como consecuencia y a raíz de mi investigación, una moralidad y preocupación por parte de la comunidad científica por el crecimiento autónomo de América Latina en estas áreas, buscando la soberanía de datos y evitando depender en exceso de infraestructuras y

financiamiento extranjeros, aunque, como ya dicho, actualmente varios auspiciadores de estas investigaciones son provenientes de Europa o Estados Unidos (Feld y Kreimer, 2019.)

Este constante vaivén entre las corrientes mundiales y la adaptación a la realidad local, los deseos y la forma única de operar en la sociedad latinoamericana, contribuye a que la región se destaque como embajadora mundial en temas como el sesgo y la ética algorítmica. Según la UNESCO (2021), hay estudios que confirman que el Sur global es más propenso a sufrir los prejuicios y sesgos relacionados con la raza, clase social, vigilancia vecinal, sistemas de justicia penitenciarios y de salud, entre otros, por lo que no es casualidad ser reconocidos como uno de los mayores difusores de la ética algorítmica en el mundo.<sup>115</sup> Esto proporciona también facilidad de tránsito entre dos mundos: analítico y práctico/ humano y tecnológico.

La implementación de esta característica también se evidencia en el área de la salud y el impacto de la epidemiología global y los estudios locales de combate y prevención de enfermedades a nivel local, por dar un ejemplo. La recopilación de datos a nivel global sobre enfermedades permite un análisis macro epidemiológico para abordar tendencias y patrones de enfermedades e índices de contagio a escala mundial. Simultáneamente, la ciencia de datos facilita investigaciones a nivel local, permitiendo un análisis específico de la propagación de enfermedades en comunidades particulares, aunque sean conducidas a partir de técnicas y saberes globalmente constituidos (Pastor Escudero, 2020).

A nivel de la esfera ambiental, vemos una relación entre la discusión alrededor de las crisis climáticas globales y su impacto a nivel regional: las investigaciones sobre el cambio climático a menudo implican análisis globales de datos climáticos (Cartea y Ángel, 2013). La capacidad de estudiar el impacto específico del cambio climático en comunidades locales, observando patrones climáticos específicos y sus consecuencias a nivel micro, aunque sus efectos se dispersen sobre todo el planeta (Letouzé, Sangokoya y Ricard, 2017).

Por fin, en el contexto de la economía, negocios y mercado, encontramos la clásica transitoriedad entre el comportamiento masivo del consumidor y el marketing personalizado. En el ámbito comercial, la ciencia de datos permite análisis macroeconómicos sobre el comportamiento del consumidor. Al tiempo en que las empresas utilizan técnicas de análisis

---

<sup>115</sup> Esa afirmación no se comprueba con un texto académico leído en los últimos años. Proviene del habla de Jair Martins de Miranda, Presidente da ABHD (Asociación Brasileña de Humanidades Digitales” en su clase Magistral de apertura del III Congreso Internacional de Humanidades Digitales” - HD Rio 2023.

de datos para personalizar estrategias de marketing, adaptándose a las preferencias individuales de clientes a nivel micro, donde la mayoría de las ventas proviene de artículos globalmente deseados y multinacionales ampliamente difundidas (González, 2020).

Como vimos, el estudio de los laboratorios de investigación social a partir de la ciencia de datos, analizados bajo las prerrogativas de los estudios constructivistas de la ciencia, nos permite trazar un interesante correlato con la personalidad contemporánea que la sociedad viene adquiriendo. Una sociedad emergida en procedimientos computacionales y en la digitalización de la vida social, lo que me llevó a llamarla y describirla como sociedad computacional. El tecno encanto, los principios de automatización y eficiencia, velocidad, identidad global y anfibia, y la alienación del trabajo por detrás de la máquina, son rasgos presentes en la vida moderna de la sociedad occidental de la que somos parte.

Defiendo, por tanto, los avances de los estudios de la CTS y sociales interesados en tecnología, pues ofrecen un interesante panorama para pensar el esquema de estructuración de la sociedad actual y de la que está por venir.

Creo que en cuanto científicos sociales estamos en un momento muy oportuno de la historia, donde nuestro oficio, saber y práctica profesional, pueden ser determinantes en el modo en que construimos nuestro conocimiento sobre lo social, y en cómo incidimos en el mundo, usando este conocimiento.

**Tabla 39.** *Resumen esquemático identidad anfibia*

Identidad Anfibia	Palabras Asociadas	Patrones Instaurados
	<ul style="list-style-type: none"> <li>* Adaptabilidad</li> <li>* Tránsito entre mundos</li> <li>* Preservación de singularidad</li> </ul>	<ul style="list-style-type: none"> <li>- Flexibilidad y adaptabilidad para moverse fluidamente entre mundos considerados distantes e incompatibles.</li> <li>- Tránsito entre “mundo natural/exactas/duro” y “mundo social/cultural/blando”</li> <li>- Preservación de la capacidad de diferenciarse</li> <li>- Construcción de puentes de comunicación</li> </ul>

A título de cierre, si bien no fueron considerados en los objetivos iniciales de este estudio, el penúltimo y siguiente capítulo es un intento de demostrar cómo los principios anteriormente mostrados y relativos a lo que llamé sociedad computacional, impacta el ejercicio de las

funciones de un profesional en ciencias sociales interesado en temas de tecnología, ciencia y sociedad.

A partir de la íntima convivencia con profesionales inmersos en este contexto, y llevando en consideración la lectura de los textos académicos analizados que indagan sobre el futuro de las ciencias sociales contemporáneas, defiendo que hay una necesidad de renovar nuestras profesiones, ya que hoy, recién vemos el inicio de nuevos cargos y necesidades profesionales que hasta hace poco no existían como tal. La nueva manera de fabricar el conocimiento social (computacional), da visión y voz a la creación de numerosas posibilidades de trabajo y actividades de investigación. Les presentaré el último hallazgo de esta investigación: el nuevo profesional de ciencias sociales.

### **Cap 10: El científico social y el futuro de la profesión: caminos para la consolidación de un saber en emergencia**

“Se o presente é digital, o futuro é analítico”

(Ricardo Capra: Pesquisador de cultura analítica, científico de datos - Brasil)

La introducción de nuevos ambientes de investigación, métodos, técnicas y artefactos, tiene un impacto significativo en la generación de nuevas demandas profesionales. En el libro *El futuro de las profesiones* (2015), los autores Richard y Daniel Susskind predicen el declive de las profesiones actuales y presentan los sistemas que las reemplazarán. Según ellos, en una sociedad mejorada por internet, no necesitaremos ni querremos que médicos, maestros, contadores, arquitectos, el clero, consultores, abogados y muchos otros, trabajen como lo hacían en el siglo XX. Explican cómo las tecnologías cada vez más capaces, desde la telepresencia hasta la inteligencia artificial, pondrán la experiencia práctica de los mejores especialistas al alcance de todos, a menudo sin o a bajo costo, y sin la necesidad imperativa de la interacción presencial (Susskind, R., y Susskind D., 2015).

En 2015, escuchar tal perspectiva puede haber sonado a ciencia ficción para muchos. Sin embargo hoy, esta realidad ya no es ni tan hipotética ni tan distante. En menos de 10 años, hemos presenciado transformaciones profundas en la sociedad, impulsadas por el avance del desarrollo tecnológico, lo que se vio acelerado por la pandemia de Covid-19 (Ponce-Tituaña y Lucio-Paredes, 2021).

Las dimensiones son masivas: miles de puestos de trabajo sustituidos por automatizaciones, drones entregando pizzas, autos que se conducen solos, robots de telemarketing de atención virtual al cliente, compras de supermercado a un clic de la pantalla del celular, son sólo algunos ejemplos de cómo ciertas actividades han hecho desaparecer ocupaciones humanas.

Al analizar este fenómeno desde una perspectiva sociológica y antropológica, en línea con los estudios constructivistas sobre ciencia, tecnología y sociedad, observamos la estrecha relación entre las tecnologías y la forma en que la sociedad se acomoda a partir de estos artefactos. A lo largo de esta investigación, hemos explorado cómo ciencia, sociedad y tecnología están interconectadas de manera interdependiente en la contemporaneidad, y cómo los cambios tecnológicos han sido capaces de transformar diversas sociedades, cada una a su manera, inclusive a nivel global.

La inclusión de la automatización en la vida, tanto en la ciencia como en la sociedad en general, ha dado origen a lo que se denominó sociedad computacional, determinada y conducida por principios como velocidad, identidad global e identidad anfibia, entre otros. A lo largo de este texto, hemos reflexionado sobre la interrelación entre el laboratorio, la ciencia y la sociedad, y en el capítulo anterior, exploramos cómo la experiencia en el laboratorio respaldó la conjetura sobre una emergente configuración de sociedad, a partir de la masificación de la digitalización de la vida y la consolidación de técnicas y tecnologías en ciencia de datos y big data.

En este capítulo me enfoco en cómo nosotros, los científicos sociales, profesionales existentes y pertenecientes a la sociedad computacional, nos vemos impactados por distintos imperativos, expectativas y capacidades que se generan de las nuevas demandas sociales. Aquí, me atrevo a ser más especulativa, ya que intento diagnosticar una demanda profesional que recién comienza a tomar forma.

A lo largo de la investigación, entrevisté a decenas de científicos involucrados en ciencia de datos y, de cuya interacción pude identificar un patrón en sus discursos: la necesidad de un tipo de profesional, de cierta formación o con ciertas habilidades específicas para esta "nueva" forma de investigar. Esto conduce a una demanda de un tipo específico de profesional en ciencias sociales, que intentaré describir en este capítulo.

A lo anterior se suma el surgimiento de una nueva disciplina en las ciencias sociales: la sociología digital, también conocida como ciencias sociales computacionales o humanidades

digitales, lo cual pude observar desde mi visión académica. Vemos, entonces, nacer un movimiento que requiere un tipo específico de profesional, al tiempo que surgen posgrados, cursos, laboratorios, redes de colaboración y asociaciones profesionales en América Latina y probablemente en todo el mundo.

Finalmente, este capítulo es un intento de demostrar de forma exploratoria, cómo la emergencia de una sociedad cada vez más conectada, globalizada, anfibia, automatizada e veloz, para dar algunos ejemplos, han provocado o instaurada la necesidad de que sean incorporados al quehacer investigativo y laboral social, cambios en la forma de se trabajar, construir y fabricar conocimientos que hablen de esta sociedad en contexto contemporáneo. La incorporación de nuevas técnicas y maneras de recolectar información a respecto de lo social , ha implicado en la necesidad de renovar nuestras profesiones. Esto se puede corroborar a partir de la observación del nacer de otras trayectorias, cargos y necesidades profesionales hasta ahora poco conocidos o desarrolladas al interior de las ciencias sociales. La introducción de la nueva forma de fabricar el conocimiento social (computacional) da lugar y voz a la creación de numerosas posibilidades de trabajo y actividades de investigación para profesionales y científicos sociales interesados o no en las temáticas de tecnología.

### **10.1 El profesional de hoy: cómo un ambiente de investigación y artefactos sociotécnicos generan una nueva demanda de profesionales**

Si bien estamos presenciando un cambio significativo en la forma de trabajar, ejercer nuestra función y talentos al servicio de la sociedad por el impacto que tecnologías computacionales, automatización e inteligencia artificial han tenido en nuestras vidas; hemos visto que este cambio también se refleja en la forma de realizar y llevar a cabo la fabricación del conocimiento científico.

Los nuevos métodos para la producción científica que hemos observado, conllevan a resultados diferentes, los cuales son generados por profesionales adaptados a nuevas demandas, formas de pensar y de ejecutar sus actividades investigativas. La interacción de los científicos sociales con este nuevo modo de investigar, conducido por tecnologías de procesamiento computacional de información, no sólo se estableció y consolidó como una alternativa viable de investigación a la tradicional, sino que también se extendió al ámbito de nuestras actividades profesionales (Needham, 2013).

La interacción con el contexto de investigaciones en ciencia de datos, así como la aplicación de este acervo de conocimiento para resolver problemas y demandas reales del universo social, abrieron las puertas para la entrada de especialistas en ciencias sociales, que a su vez, se transformaron y contribuyeron a modificar la forma en que trabajaban (Guba, 2018).

Mi argumento es que esta identidad híbrida, generada desde el sistema de investigación social a partir de la ciencia de datos y la interdisciplinariedad de profesionales enfocados en un objetivo aplicado, está cambiando la dinámica de la producción científica en variadas áreas del conocimiento. Este análisis surge de la observación de una constante queja sobre la falta de habilidades humanas adecuadas para liderar, participar y colaborar en proyectos de investigación aplicada con enfoque en ciencia de datos. En este estudio, se constata desde tres instancias: en las más de 50 entrevistas a especialistas en ciencias sociales y otras disciplinas, a través de mi participación directa en diferentes contextos de investigación de las ciencias sociales digitales en Brasil y Chile, y por último, en los numerosos eventos, seminarios y cursos que he presenciado a lo largo de estos cinco años de investigación.

Sin embargo, antes de detallar las tres principales características, rasgos y actividades del nuevo profesional en ciencias sociales, permítanme realizar una breve digresión para presentar de manera más resumida el tipo de profesional actualmente involucrado en estas investigaciones en curso: el no involucramiento directo con los datos, la capacidad de traducción entre dos mundos, y asumir funciones de liderazgo operacional y de gobernanza de proyecto son los principales atributo y/o funciones que han desarrollado los profesionales en ciencias sociales en la actualidad.

**Tabla 40.** *Principales rasgos del nuevo profesional en ciencias sociales*

Características actuales del profesional en ciencias sociales	Descripción
No se involucran directamente con los datos	<ul style="list-style-type: none"> <li>● No involucrarse directamente con los datos.</li> <li>● Argumentan que trabajar en colaboración es más eficiente que dedicarse el tiempo de aprender todos los detalles técnicos.</li> <li>● Realizan trabajos tradicionales de las ciencias sociales (surveys, entrevistas, etnografía, grupo focales) dentro de dicho universo de investigación</li> <li>● Actúan como analistas o consultores en sus temáticas de expertises.</li> <li>● Validan los resultados y auxilian en el proceso de mejoramiento de la database de entrada</li> </ul>

<p><b>Ser un puente entre dos mundos</b></p>	<ul style="list-style-type: none"> <li>● Enlace entre el conocimiento técnico y las necesidades humanas.</li> <li>● Traductores entre las necesidades de los contratantes o demandas del proyecto y la conversación entre las posibilidades reales de ejecución del proyecto con los desarrolladores y científicos de datos</li> <li>● Son el punto de contacto directo con el servicio, interactúan con los clientes y usuarios del proyecto.</li> <li>● actúan como "hacedores de puentes", comunicándose entre dos mundos : identidad anfibia a florada</li> <li>● Mediadores entre saberes técnicos computacionales y saberes sociales</li> <li>● Inseminan la importancia de medir y evaluar sesgos, discriminaciones y mal uso de las informaciones trabajadas</li> <li>● Puente entre saberes prácticos y saberes teóricos. Involucramiento de la academia</li> </ul>
<p><b>Funciones de liderazgo operacional y de gobernanza</b></p>	<ul style="list-style-type: none"> <li>● Ejercen funciones de liderazgo operacional y gobernanza en el proyecto.</li> <li>● Se encargan de funciones más administrativas: encontrar presupuestos, redactar proyectos, gestionar relaciones públicas, redactar informes, evaluar el proceso</li> <li>● Interconexión entre tres entidades: Organizaciones gubernamentales y instituciones de desarrollo desenvolvimiento científicos y social economico + Tercer sector: empresas desarrolladoras de sistemas, start ups, empresas consultoras en datos o empresas multinacionales que venden las horas de procesamiento de datos o almacenamiento seguro de la información + Academia: colaborando con desarrollo de novos modelos, tecnologías de investigación, e alto cuadro de profesionales altamente cualificados</li> </ul>

Con respecto a la afirmación de que no se involucran directo con los datos, se complementa con los datos recopilados a través de las entrevistas con expertos en la materia y un estudio etnográfico realizado entre 2021 y 2022, donde se observa que la mayoría de estos profesionales se involucran en proyectos de colaboración externa, desempeñando roles subordinados (Kreimer, 2019) o realizan actividades puntuales dentro del contexto general de investigación. En otras palabras, su participación no se traduce comúnmente en roles de dirección, coordinación o investigación principal en dichos proyectos y en condiciones muy semejantes a las que Adriana Feld y Pablo Kreimer (2019) describen en sus estudios sobre los científicos latinoamericanos investigando por el mundo.

Un hallazgo relevante es que, en su mayoría, los investigadores utilizan tecnologías big data y procesamiento computacional de la base de datos de manera tangencial, complementando o informando procesos investigativos específicos. La mayoría de ellos prefieren no involucrarse directamente con los datos, argumentando que trabajar en colaboración es más eficiente que dedicarse al tiempo y esfuerzo necesario para aprender todos los detalles técnicos que involucran el procesamiento computacional de datos.

Aunque es cierto que la mayoría de los entrevistados científicos en ciencia sociales consideran que es más eficaz trabajar en conjunto a un especialista, que tener que aprender a

hacer ellos mismos este trabajo, me parece importante dejar registro de que este pensamiento no es del todo un consenso. Hay una fuerte controversia al respecto, sobre todo entre los profesionales más jóvenes. Por lo general, los científicos sociales más jóvenes, de un máximo de 28 años, afirman que es deber del científico social del presente estar en contacto, aprender e incluso ser capaces de programar, o ejecutar análisis de datos a través de algún sistema de estos, como el R. o Python. Consideran que este saber puede ser fácilmente aprendido como cualquier otra disciplina que se estudia en las carreras en ciencias sociales.

Por otro lado, los científicos más avanzados de edad, e incluso, infelizmente me encuentro entre estos académicos, consideran ser más eficiente invertir en un buen y fluido proceso de comunicación e investigación cooperando con estos profesionales, a fin de alcanzar con mayor rapidez y mejor calidad los resultados requeridos. Lo que garantizaría, inclusive, una cierta distancia con el procesamiento de los datos, cosa que para este contexto de investigación, puede ser considerado una ventaja, en el sentido de estar más alejado, ser capaz de mirar los errores y aportar con sugerencias, ideas y opiniones con más distancia respecto al mundo computacional.

Según la opinión de varios, la introducción de la diferencia en la investigación es originada a través del pensamiento discordante, y lo que acaba agregando valor a la investigación en sí, es la propia función del especialista en ciencias sociales. Estos profesionales desempeñan, entonces, roles como analistas de datos, o como consultores en sus áreas de especialización. Entregan un conocimiento significativo desde las ciencias sociales, contribuyendo al análisis e interpretación de los datos generados en proyectos de esta naturaleza.

No obstante, este distanciamiento operacional no impide que ellos, inmersos en dicho sistema de investigación y a través de colaboraciones entre los diferentes especialistas, reflexionen y teoricen sobre las complejidades del sistema socio-técnico que rodea a la ciencia de datos y big data. La interconexión entre áreas "más humanas" y conocimientos "más tecnológicos", considerados tradicionalmente antagónicos, se manifiestan en este sistema donde ambas perspectivas convergen, generando una co-dependencia en la producción de conocimiento, ya sea aplicado o teórico.

Estos profesionales continúan realizando labores tradicionales de las ciencias sociales, adaptándolas al contexto de la investigación basada en ciencia de datos. Sus habilidades en la realización de encuestas, entrevistas, etnografía y grupos focales aportan una perspectiva

valiosa al proceso, ya que entregan una mirada más profunda sobre los diversos procesos que permean la temática de la investigación. Lo que hace destacar la importancia de la introducción de diferentes especializaciones en el contexto de esta investigación, la utilización de técnicas mixtas de producción de información y el análisis del conocimiento adquirido y de la colaboración interdisciplinaria en proyectos de esta naturaleza.

Después de algún tiempo y con un poco más de experiencia, empecé a preguntarme sobre la conformación de mi oficina: ¿Mi oficina de datos, mis analistas, saben hacer una entrevista como la que tú estás haciendo aquí ahora, pensando en su trabajo? Importante también sería vincular todo esto al background de los modelos predictivos y estadísticos y a las demás metodologías que utilizamos. ¿Realmente estamos consiguiendo añadir a estas cajas opacas de la ciencia de datos toda la inteligencia del caso de estudio y la cercanía a los desafíos reales de los problemas que intentamos resolver!? Mira, no es fácil, pero lo estamos intentando, tenemos que ser capaces de hacer que estas dos cosas hablen. (WLM - Entrevista 12/2023 - Brasil . mi traducción)

Además, las entrevistas revelan inquietudes éticas y epistemológicas sobre la producción de conocimiento en el contexto de la ciencia de datos y de la computación. Se evidencia una migración de profesionales de tecnología de la información hacia la ciencia de datos, pero muchas veces sin una reflexión sobre los límites éticos y epistemológicos que este traslado disciplinario implican en ambos saberes y al interior de las propias disciplinas. Se destaca la preocupación por la falta de comprensión en algunos entornos de la importancia de combinar modelos predictivos y estadísticos con la habilidad de realizar documentación del proceso investigativo, lo que plantea cuestionamientos sobre la transparencia y el entendimiento de las operaciones en el ámbito de las ciencias de datos.

Estas percepciones iniciales proporcionan un punto de partida sólido para explorar más a fondo las dinámicas de la participación de los científicos sociales en proyectos de big data y señalan la necesidad de considerar de manera crítica los desafíos éticos, metodológicos y epistemológicos en este espacio emergente.

El mayor desafío actualmente vivenciado por estos profesionales, es lo que tiene que ver con los desafíos de aprendizaje tecnológico: ser capaces de comprender el vocabulario estándar

de las ciencias de la computación. Estar familiarizado con las soluciones posibles, conocer mínimamente el mecanismo de funcionamiento de un laboratorio de vocación computacional. Lo que nos lleva directamente al segundo aspecto, o característica presente el científico social del presente, en el contexto de investigación computacional.

La traducción entre dos mundo, es decir, la comunicación entre las áreas y saberes “más humanos/ blandos” con aquellos conocimientos “más tecnológicos/exactos/duros”, tradicionalmente consideradas antagónicas por la comunidad científica. En el contexto del sistema de investigación estudiado, acaba por instalarse en el abanico de habilidades de estos profesionales, una comprensión por la comunicación muy destacada.

Como los profesionales actúan como enlace entre el conocimiento técnico, representado por expertos en ciencia de datos, y las necesidades humanas, asegurando una comprensión mutua y facilitando la colaboración efectiva, acaban volviéndose los traductores de las necesidades de los contratantes o de las demandas del proyecto, abogando por la conversación entre las posibilidades reales de ejecución de las soluciones con los desarrolladores y científicos de datos.

Actúan, además, como el punto de contacto directo con el servicio prestado a través de las soluciones o resultados generados, interactuando con los clientes y usuarios del proyecto. Esta interacción directa garantiza una comprensión clara de las expectativas, y permite ajustar la ejecución del proyecto según sea necesario, desempeñando con más primor el “carácter anfibio” entre las distintas áreas, presentado con más detalle en el capítulo anterior.

Esta habilidad para navegar y comunicarse efectivamente en ambos contextos es esencial para el éxito de la colaboración entre disciplinas, su rol de mediadores entre saberes técnicos computacionales y sociales garantiza que los aspectos técnicos del proyecto se comprendan en un contexto social más amplio. También, viceversa, se aseguran que la contraparte técnica del proyecto sea capaz de acceder y conocer lo necesario a las temáticas trabajadas, logrando ajustar de la mejor manera posible su trabajo a las necesidades reales y específicas de la sociedad a la que se dirige el proyecto.

Es justamente la mirada más distante con resto del proceso de tratamiento de datos, que logra diseminar la importancia de medir y evaluar sesgos, discriminaciones y mal uso de la información trabajada, destacan la importancia de medir y evaluar sesgos, discriminaciones y

el posible mal uso de la información, promoviendo la ética en la investigación y mitigando posibles impactos negativos.

Esta es la manera como entregan y añaden valor al proyecto al ser el enlace entre la complejidad técnica y la comprensión humana. ¿Cómo se traduce esta "traducción" en la creación de soluciones más efectivas y socialmente relevantes? Al impulsar este tipo de debate, los compromisos éticos se materializan en una comunicación más transparente y que va más allá de la responsabilidad de comunicar claramente cada paso del proceso. La presentación cuidadosa del camino hacia las conclusiones, junto con la delimitación precisa de las capacidades y limitaciones, se convierte en una parte integral de la narrativa investigativa en el sistema socio técnico de las investigaciones sociales a través de la ciencia de datos.

Por fin, la última característica o función ejercida por los profesionales en ciencias sociales involucrados en las investigaciones en ciencia de datos, son las funciones de liderazgo operacional y gobernanza en el proyecto: guiando la ejecución del mismo y asegurando la alineación con los objetivos establecidos. Asumen responsabilidades administrativas cruciales, incluyendo la identificación de presupuestos, redacción de proyectos, gestión de relaciones públicas, elaboración de informes y evaluación continua del proceso. Su enfoque administrativo es esencial para el flujo eficiente de la investigación.

Otro elemento importante y bastante presente en estos profesionales es la interconexión estratégica entre tres entidades: Organizaciones gubernamentales y diferentes servicios públicos; instituciones, ONGs e institutos de desarrollo científico; y socioeconómico, como el tercer sector, que abarca empresas desarrolladoras de sistemas, *startups*, consultoras de datos o empresas multinacionales que ofrecen servicios de procesamiento de datos y almacenamiento seguro de información; y la academia, colaborando en el desarrollo de nuevos modelos, tecnologías de investigación y contando con un alto cuadro de profesionales altamente cualificados. Esta interconexión amplía el alcance y la pertinencia de los proyectos.

Acá cabe resaltar que respecto al contexto latinoamericano, se destaca el desarrollo de una capacidad específica y necesaria para poder enfrentar los desafíos comunes en la interacción con los organismos públicos. Desde vencer procesos burocráticos hasta adaptarse a entornos no muy innovadores, se aborda el impacto de las reglas gubernamentales en el uso del tiempo, la asignación de recursos y la estructura organizativa del estado. La capacidad de

navegar eficazmente estos desafíos se convierte en una competencia clave para estos líderes operacionales en proyectos de este tipo.

Esta etapa manual, no automatizada de la investigación, implica un esfuerzo considerable, en cuanto a tiempo y profundidad de análisis. Además, desempeñan un papel crucial en la curaduría de datos, llevando a cabo la limpieza del corpus de datos, eliminando información innecesaria y orientando el proceso de aseguramiento de la calidad de los datos. Esta meticulosa labor contribuye a la robustez y confiabilidad de los análisis, destacando la importancia de su participación en cada fase del proyecto.

Por último, con respecto específicamente a la dimensión de formación de estos profesionales, la gran mayoría de los casos analizados de los entrevistados y también de la base de datos, poseen por lo menos un título de especialización, magíster. El 80% de los entrevistados son doctores. Este hallazgo se torna más interesante, desde mi punto de vista, porque siendo todos ellos postgraduados, casi todos doctores, no exploran o no poseen visibilidad en el mundo de los artículos académicos. No circulan conocimiento preferencialmente por esta vía. No porque no produzcan, pero usan otros formatos como capítulos de libros, guías y normativas de utilización masiva, informes de proyectos de investigación aplicada, resúmenes de conferencias y anuarios de congreso.

En concreto, el profesional contemporáneo en ciencias sociales inmerso en la investigación basada en datos, se destaca por su especialización en el análisis de datos y la habilidad para traducir estos análisis en acciones y actividades concretas, principalmente en contextos vinculados a iniciativas gubernamentales o públicas. Su expertise radica en la toma de decisiones respecto a los resultados, una fase manual e intensamente analítica de la investigación. En este sentido, participa activamente en el establecimiento de un diálogo con colegas y expertos técnicos, definiendo qué aspectos serán contemplados en el análisis y cómo se transformarán en decisiones prácticas.

Con estas características delineadas, se abre la puerta a explorar las futuras actividades y roles de estos profesionales en el ámbito de la ciencia de datos aplicada a las ciencias sociales. A continuación, presentaremos las actividades emergentes y las competencias necesarias para enfrentar los desafíos de este cambiante escenario profesional.

## **10.2 El profesional del mañana y la consolidación de las ciencias sociales computacionales**

Primero se discutió sobre las transformaciones profesionales que vienen ocurriendo en el ejercicio de las investigaciones en ciencias sociales y las implicaciones metodológicas que han modificado el proceso de la producción social, y, en consecuencia, de la sociedad computacional. Después se presentó el perfil de los profesionales en ciencias sociales involucrados en el sistema de investigación social en ciencia de datos y como tales comportamientos se relacionan con la emergencia de un nuevo perfil de trabajo en la sociedad que está emergiendo. Ahora, les presentaré una especulación en cuanto a los profesionales en ciencias sociales demandados en un presente/futuro que considero estar próximo.

La idea de este "profesional del futuro" es una conclusión, como mencioné, de escuchar y convivir en el entorno de mi investigación. Fueron surgiendo a partir de preguntas del tipo: ¿Qué habilidades, según tu opinión, son las que más hacen falta en los profesionales involucrados en proyectos en ciencia de datos?, ¿En que más te fijas a la hora de contratar o invitar a participar a un especialista en ciencias sociales? ¿Cuál es el tipo de saber que más hace falta en estos profesionales? Si tuvieras todos los recursos del mundo y todos los profesionales disponibles para tu proyecto, ¿qué tipo de investigación te gustaría llevar? ¿Qué tipos de problemas o proyectos implementarías y por cual relevancia? ¿Cómo serían las características de los profesionales en ciencias sociales que te gustaría incorporar a tus proyectos?

Preguntas como estas posibilitaron recorrer un tipo de información que habla respecto a un profesional en ciencias sociales escaso en el presente, y muy deseado en un futuro no tan distante. Son consecuencia de la intersección de las ciencias computacionales, las tecnologías big data, y el procesamiento y automatización de la vida. Características que han redefinido nuestro entorno, el mundo del trabajo, y evidentemente, del oficio y las características de muchos de los científicos sociales.

Este profesional desempeña un papel esencial en la conformación e hilo conductor de las dinámicas sociotecnológicas de la contemporaneidad. Como un clásico etnógrafo, son solicitados con la finalidad de tejer puentes entre distintos saberes, hablar el idioma de la gente (el usuario o beneficiario de los sistemas y soluciones fabricadas), mediar el interés

entre los poderes público y privado, y finalmente, contar al mundo lo que han aprendido sobre nosotros mismos en cuanto sociedad. Pero para más allá de la práctica de la antigua y tradicional labor socio-antropológica, su quehacer se ha extendido más allá de los límites tradicionales de las actividades de la profesión.

Ya no es suficientes el lápiz y papel, ni siquiera un computador y un programa editor de texto. La artesanía del oficio abarca un conjunto de actividades cruciales para asegurar la coexistencia armoniosa entre la tecnología y la sociedad. Este nuevo profesional se erige como un facilitador fundamental en la participación ciudadana, particularmente en el diseño de proyectos de incidencia social y de tecnologías asociadas a los programas gubernamentales, y la automatización de sus sistemas. Además, se adentra en terrenos complejos, como el análisis de riesgos en temas ambientales, de seguridad pública y distribución energética, por dar algunos ejemplos. Están directamente involucrados en proyectos, y construcción de medidas de evaluación de impactos, estándares de resolución de conflictos globales, y la elaboración de mecanismos de remediación para posibles daños.

Como pioneros, o vanguardia de este conocimiento, la creación de marcos normativos, guías y sellos de calidad, se convierte en parte intrínseca de su labor. Tales materiales trazan camino para garantizar la integridad y ética en el uso de tecnologías, y el bien vivir de la sociedad computacional. Como extensión de sus roles, son importantes referentes y asesores en temas de regulación, legalidad y protección datos, constituyendo así para la creación y regulación de sistemas y sus buenos usos en un entorno cada vez más digital y complejo.

A través de núcleos de formación profesional, los científicos sociales computacionales, también cultivan el conocimiento y las habilidades necesarias para formar, amateur, ingenieros y/o programadores en materia de investigación y metodología social, en los temas de transparencia de datos, ciudadanía y buen uso de la infraestructura computacional para el tratamiento de datos sociales. De este modo, lideran la mediación y capacitación de ingenieros y tecnólogos humanistas: son expertos en el arte de la traducción entre el lenguaje humano y la maquinaria digital.

Cuando yo estaba en el último año del colegio, en la etapa de elegir una profesión para estudiar en la universidad, nuestra orientadora de carrera hizo circular entre los estudiantes una “guía de las profesiones de 2006”, una especie de catálogo que listaba los diversos tipos de carrera, con las respectivas universidades que la enseñaba, el sueldo mínimo medio de

entrada, una breve descripción de las disciplinas que involucran el aprendizaje de tal profesión, y la descripción de lo que uno podría hacer, después de recibido el título.

Casi 20 años después, acá estoy, ejerciendo la profesión que elegí a través de la lectura de muchos manuales hechos para preuniversitarios y escribiendo un tipo similar de catálogo de profesiones. Muchas veces me pego pensando en cuál sería el contenido y la definición de las “ciencias sociales” en las guías para estudiantes del año de 2023. ¿Qué habrá cambiado en el contenido de esta definición de un tiempo a esta parte?

La tabla a continuación pretende ilustrar algunas actividades solicitadas o descritas como necesarias por los entrevistados de este estudio. Resume la descripción de las funciones o actividades requeridas para cada función identificada durante mi investigación, acompañada de un ejemplo de implementación vinculada a determinadas áreas temáticas.

**Tabla 41.** *Listado de actividades del nuevo científico social computacional*

Actividad	Descripción
<b>1- Facilitación de procesos de participación en el diseño de políticas y tecnologías asociadas</b>	<p>Mediación activa para fomentar la colaboración entre ciudadanos, usuarios y diseñadores de políticas basadas en algoritmos de decisión, garantizando la inclusión de diversas voces en el diseño de sistemas automatizados.</p> <p>Ejemplo: Intermediación de un proyecto nacional de implementación de un sistema de clave única, o entrega de un subsidio social, como es el caso de la designación de los niños a los colegios públicos, o la designación de viviendas sociales, etc.</p>
<b>2- Análisis de riesgos y evaluación de impactos decorrentes de tecnologías computacionales y automaciones</b>	<p>Identificación proactiva de posibles daños y evaluaciones exhaustivas de impactos sociales de políticas y tecnologías, contribuyendo a estrategias de mitigación temprana.</p> <p>Ejemplo: Estrategias y planos de identificación de sesgo de las datas de entrada y de salida de los proyectos; sensibilización e introducción de pautas de ética de datos. Participación activa en el proceso de calidad de los datos.</p>
<b>3- Mecanismos de remediación de daños y perjuicios</b>	<p>Desarrollo de estrategias efectivas para remediar daños y perjuicios identificados, inspiradas en prácticas socioambientales, socioeducativas y medidas paliativas que corrigen y compensan perjuicios.</p> <p>Ejemplo: ¿Qué pasa cuando una persona es identificada erróneamente a través de un sistema de reconocimiento facial de un programa de seguridad pública como responsable por cometer un delito?</p>
<b>4- Creación de marcos normativos, guías y sellos de calidad</b>	<p>Contribución a la creación de marcos legales, guías y sellos de calidad que aseguren la transparencia, equidad y eficacia de las políticas y tecnologías implementadas.</p>

	<p>Ejemplo: Construcción de guías y manuales tanto para desarrolladores que quieren introducir parámetros éticos en sus trabajos, como a equipos que desean implementar técnicas computacionales a investigaciones sociales, tanto para gestores y administradores públicos que quieran iniciar proyectos en sus servicios, etc...</p>
<b>5- Asesoría en marcos legales para protección de datos y usos de datos</b>	<p>Trabajo en pro de la aplicación de marcos legales específicos a la realidad de cada proyecto (si público, si social, se empresarial o para fines de la ciencia), asegurando la protección de datos y regulando el desarrollo y uso ético de sistemas tecnológicos a contextos diversos.</p> <p>Ejemplo: Implementación de un sistema de seguridad de datos, gestión y gobernanza de datos en proyectos cuyas expertises tangencian los proyectos.</p>
<b>6- Núcleos de formación profesional</b>	<p>Liderazgo en la formación de “profesionales anfibios” a través de diplomados, maestrías y talleres, incluyendo la capacitación de ingenieros con sensibilidad social y la incorporación de humanistas.</p> <p>Ejemplo: cursos y diplomados de ética de datos, procesamiento de datos públicos, ciencia de datos y políticas públicas, etc. Por otro lado, ofrecer formación a profesionales en humanidades, capacitando en programación, o metodologías digitales.</p>
<b>7- Estudios de opinión masiva entre usuarios y plataformas</b>	<p>Conducción de estudios de opinión para alinear tecnologías con expectativas y valores de la sociedad, actuando como puente entre las necesidades de los usuarios y las capacidades de las plataformas.</p> <p>Ejemplo: Estudios de impacto de utilización de las herramientas; aperfeicoamento y optimización de trabajo, cuando las soluciones son de gestión interna a una organización, usabilidad de plataformas de servicios y canal de atención al ciudadano; Construcción de plataformas de consulta pública; Estudio de percepción política a escala global.</p>
<b>8- Estudios y mapeamento de stakeholders y redes de influencia a partir de la minería de datos</b>	<p>Analizar a las partes interesadas clave (stakeholders) y las conexiones de influencia en un determinado contexto. Esto incluye la extracción, procesamiento y análisis de datos relevantes para comprender las relaciones entre diferentes actores, sus intereses y su impacto en un sistema o proyecto específico</p> <p>Ejemplo: mapear los principales nombres de los estudios de legislación y protección de datos de un país, con la finalidad de armar un red de profesionales capaz de impulsar determinadas agendas en los diferentes procesos legislativos en cada país o región.</p>
<b>9- Etnografía en ciencia de datos</b>	<p>Aplicación de métodos etnográficos en el contexto de la ciencia de datos y el procesamiento computacional de información. Este enfoque combina la observación participante, entrevistas y análisis cualitativo para comprender en profundidad los aspectos sociales, culturales y éticos asociados a la implementación de proyectos de ciencia de datos. El etnógrafo en ciencia de datos se sumerge en los entornos de trabajo, colabora con equipos multidisciplinarios y analiza las interacciones humanas con los sistemas tecnológicos</p>
<b>10- Participación y consultoría en Proyectos de ciencia de datos</b>	<p>Integración activa entre su expertise temática y de trabajo del proyecto, los profesionales en el diseño, desarrollo y ejecución de proyectos de ciencia de datos, actuando como consultores especializados en aspectos sociales y éticos.</p>

<p><b>11- Implementación de escritorios de datos en grande institución y diferentes servicios del gobierno</b></p>	<p>Implica liderar el diseño, desarrollo e implementación de plataformas de datos a gran escala. Estos profesionales desempeñan un papel crucial en la creación de infraestructuras que facilitan el acceso, análisis y gestión eficiente de grandes volúmenes de datos. Además, se centran en garantizar la seguridad, integridad y ética en el manejo de la información.</p> <p>Ejemplo: Su labor incluye colaborar con equipos multidisciplinarios, establecer políticas y procedimientos para la gestión de datos, y asegurar la interoperabilidad entre diferentes sistemas y servicios gubernamentales. La implementación de estos escritorios de datos busca mejorar la toma de decisiones, la eficiencia operativa y la transparencia en la administración pública.</p>
<p><b>12- Gerente de proyecto en ciencia de datos</b></p>	<p>Planificación, ejecución y supervisión de iniciativas relacionadas con la ciencia de datos. Este profesional lidera equipos multidisciplinarios, colaborando estrechamente con científicos de datos, ingenieros y otros especialistas. Su responsabilidad incluye la definición de objetivos del proyecto, asignación de recursos, gestión del cronograma y aseguramiento de la calidad. Además, desempeña un papel clave en la comunicación efectiva con los interesados, asegurando la alineación de los resultados del proyecto con los objetivos organizativos y estratégicos.</p>
<p><b>13- Analistas de datos</b></p>	<p>Interpretación, procesamiento y análisis de conjuntos de datos utilizando herramientas y técnicas avanzadas de ciencias de datos. El analista de datos se especializa en la transformación de datos crudos en información significativa, extrayendo patrones, tendencias y conocimientos que puedan respaldar la toma de decisiones informadas. basado en sus respectivas expertises</p>

Esta descripción destaca el papel esencial de los profesionales actuales en ciencias sociales, considera cómo la inserción en la sociedad computacional está dando forma a la evolución del campo profesional de las ciencias sociales. Como se puede notar, los científicos sociales computacionales actúan como un mediador clave, fomentando la colaboración de las diversas y anfibias voces de la sociedad, influyendo y promoviendo la consideración de la introducción de aspectos computacionales en la vida social.

Las actividades mencionadas en la tabla 40 pueden ser reagrupadas y clasificadas a través de tres grandes grupos de funciones: especialistas, de formación y analista.

### ***10.2.1 Los especialistas: expertos interdisciplinarios integrados en atribuciones emergentes***

Ejercen una infinidad de funciones en situaciones donde son requeridos como expertos en un tema específico, pueden actuar como responsable de la fase inicial de los proyectos de

investigación, donde, por ejemplo, seleccionan y determinan las condiciones y principales variables y categorías a ser incorporadas al proyecto. Implica la integración activa de profesionales en el diseño, desarrollo y ejecución de proyectos de ciencia de datos, actuando como consultores especializados en aspectos sociales y éticos. Estos profesionales participan en la definición de objetivos y en la evaluación final de las soluciones y resultados. Además, ofrecen asesoramiento para garantizar la transparencia, equidad y responsabilidad en el uso de datos.

También pueden actuar en la fase de validación de los datos, no solamente como analistas, sino que también como identificadores de daños o perjuicios sociales con respecto a la implementación de tecnologías automáticas, del procesamiento computacional de la información y la inteligencia artificial. En este rol, el profesional propone y desarrolla mecanismos efectivos de remediación, trabajando en la implementación de medidas paliativas que buscan corregir y compensar los posibles prejuicios generados por la tecnología. Además de identificar y mitigar posibles daños, este profesional lleva a cabo evaluaciones exhaustivas de los impactos de las políticas y tecnologías propuestas. A depender de su experiencia en este tipo de actividad, actúa de forma bastante importante en pro de la identificación temprana de riesgos, lo que por consecuencia auxilia en el proceso de garantizar la seguridad, responsabilidad, transparencia y ética en la implementación de nuevas soluciones en proyectos o iniciativas posteriores.

Otro tipo de función, que puede verse con bastante presencia, es la implementación de escritorios de datos en grandes instituciones y/o diferentes servicios del gobierno, lo que implica liderar el diseño, desarrollo e implementación de plataformas de datos a gran escala dentro de instituciones gubernamentales y diversos servicios públicos, o bien en ciertos sectores de grandes empresas. Estos profesionales desempeñan un papel crucial en la creación de infraestructuras que facilitan el acceso, análisis y gestión eficiente de grandes volúmenes de datos. Además, se centran en garantizar, a través de su asesoría y experticia, la fomentación de un plan de gobernanza de datos que garantice y mantenga la seguridad, integridad y ética en el manejo de la información pública en el Estado.

En este contexto, la actividad también puede asumir la responsabilidad en la gestión de procesos, la selección de profesionales especializados a ser estratégicamente incorporados en estos ambientes, contribuyendo para la reorganización logística de equipos, funciones y personas. Los encargados de la implementación no sólo pueden liderar en términos técnicos,

sino que también, su actividades puede incluir el desafío de instaurar una cultura de colaboración multidisciplinarios, al establecer procedimientos en gestión de datos en espacios investigativos, se tornando así el responsable por coordinar la interoperabilidad entre las capacidades de infraestructura y gobernanza de datos, siempre visando la eficiencia operativa de los proyectos y su estabilidad funcional.

### ***10.2.2 Formación: la importancia de multiplicar un campo emergente***

El segundo grupo de funciones desempeñadas por los científicos sociales computacionales habla del liderazgo con respecto a la formación de profesionales híbridos, por medio del establecimiento de núcleos educativos, centros de formación especializada, diplomados, maestrías y talleres. Su enfoque incluye la capacitación de ingenieros y programadores con sensibilidad social, o la llamada incorporación de *soft skills* (Ribes, 2018), así como la incorporación de humanistas en el lenguaje y la traducción entre los ámbitos humano y tecnológico o *hard skills*, según Ribes (2018). Actúan como un puente entre las necesidades de los usuarios y las capacidades de las plataformas, lleva a cabo estudios de opinión exhaustivos con el objetivo de asegurar la alineación de tecnologías con las expectativas y valores de la sociedad.

También pueden desempeñar el papel de educadores de la sociedad civil, a través de la incorporación de sus trabajos en ONGs, instituciones y organizaciones internacionales, cursos y universidades libres, abiertas de modalidad a distancia. Sus conocimientos pueden ser utilizados como herramienta de transformación social y sensibilización de la población sobre las implicación y consecuencias del vivir en una sociedad computacional. Educando al público sobre temas importantes, se podría influir en la toma de decisiones y comportamientos cotidianos del ciudadano medio.

Desempeñándose como una suerte de arquitecto normativo, otro papel de este profesional con enfoque en la formación, contribuye a la creación de marcos legales, guías y sellos de calidad que garantizan la transparencia, equidad y eficacia de las políticas y tecnologías implementadas. Su labor sienta las bases en el sistema de ley de protección de datos, guiando e implementando la creación y aplicación de marcos legales específicos, para un entorno ético y confiable, sea en el Estado (a través de asesorías y en el ejercicio de sus cargos

públicos), es decir, en proyectos de investigación aplicada o incluso en empresas o diferentes tipos de escritorios de tecnología.

Asegurando la difusión e implementación de protección de datos, regulación y desarrollo de sistemas, su perspicacia jurídica se aplica a una variedad de contextos: desde su mayor incidencia en proyectos relacionados al derecho y justicia, temas regulatorios y legislativos, pasando por la salud y telemedicina, psicología y estudios de comportamiento, neuropsicología, prevención a violencia psicológica y abusos; economía y la predicción de fraudes al fisco, predicción y proyección económica, consumo; administración pública y la utilización de servicios, digitalización del estado, distribución de gastos y recurso públicos.

Las actividades trabajadas al interior de las actividades de formación, como vimos, pueden ser extremadamente variadas y estar vinculados a distintos enfoques. He observado a lo largo de la investigación la contribución de diversos profesionales en ciencias sociales apoyando el proceso de creación y conformación de asociaciones en humanidades digitales, redes de académicos y trabajadores interesados en inteligencia artificial, colaboración en la instauración de posgrados sobre ciencias de datos y datos sociales en universidades chilenas y brasileñas, y también acompañé una reunión donde la decana de una universidad en Chile, reclutó a algunos de los más destacados profesionales en el país, para hacer una propuesta de creación de una graduación en ciencias sociales digitales en su universidad. Es nítido el interés de parte de varias instituciones en formalizar este tipo de conocimiento que recién emerge en la academia y en el campo de las humanidades, como veremos en el último apartado de este capítulo.

### ***10.2.3 Los Analistas: el ejercicio de la “función humana” en contextos tecno-encantados***

Este tercer grupo de funciones está relacionado con la labor de identificar prácticas, percepciones y dinámicas sociales que pueden influir en la implementación y en los resultados de proyectos de ciencia de datos, contribuyendo a un enfoque más holístico y socialmente consciente en este campo. La actividad de analizar datos, en el contexto de las ciencias sociales y el procesamiento computacional de información, implica la interpretación del conjunto de datos utilizando herramientas y técnicas avanzadas de ciencias de datos.

El analista de datos se especializa en la transformación de datos crudos en información significativa, extrayendo patrones, tendencias y conocimientos que puedan respaldar la toma de decisiones informadas. En el ámbito de las ciencias sociales y el procesamiento computacional de información, el analista de datos juega un papel crucial al aplicar métodos estadísticos y algoritmos para comprender fenómenos sociales, evaluar impactos y contribuir a la generación de conocimiento científico. Además, el analista de datos puede colaborar estrechamente con científicos sociales, desarrolladores y otros profesionales para garantizar la calidad y relevancia de los resultados obtenidos.

Creo que esta función es una de las más valoradas actualmente. Justamente por la escasez de profesionales capaces de poseer tales habilidades técnicas e incluso de bagaje teórico y tecnológico suficiente. También es un rol bastante variado, porque puede englobar una infinidad de actividades, en proporción de la cantidad de problemáticas y especificidades sociales que puedan existir.

Se vuelve especialmente delicado analizar tantas y diversas partes interesadas, actores, artefactos y conexiones claves de influencia en un contexto que recién emerge, y por lo tanto, que no es tan estable. Esto incluye la extracción, procesamiento y análisis de datos relevantes para comprender las relaciones entre los diferentes actores, sus intereses y su impacto en un sistema o proyecto específico, permitiendo una visión más completa y estratégica de las dinámicas de influencia en un entorno de análisis determinado.

Trabajar como analista en un contexto de investigación social a partir de la ciencia de datos significa volver a trabajar con el análisis de grandes conjuntos de datos relacionados con temas sociales y políticos globales o continentales. Gracias a las metodologías y técnicas de producción de información de redes sociales, se puede implementar minería de texto volcados a datos sociales, análisis espaciales y geolocalizadas para la comprensión de territorios y ocupaciones en grande y pequeña escala; técnicas de análisis de datos longitudinales; estudios de mercado y consumo a través de internet, por nombrar algunos.

Los analistas socio-computacionales de datos están inmersos en un entorno tecnológico avanzado, lo que implica estar familiarizado con las herramientas y plataformas de ciencia de datos adecuadas a cada objetivo del proyectos en cuestión. Aprovechan este saber, el poder del aprendizaje automático y análisis de grandes datos para así extraer entendimientos diferenciados a la luz de otros procesos de producción del conocimiento. Por trabajar con una

diversidad mayor de datos y en tiempo real, la capacidad para procesar, analizar y derivar conclusiones a partir de la colaboración de sistemas de análisis de grandes conjuntos, es esencial en este contexto. Este conocimiento también es adquirido en la práctica, que recién se está consolidando en el mundo del trabajo en humanidades, haciendo que este profesional sea aún más requerido.

Otro elemento que hace que el analista de datos actual se destaque, es su capacidad de trabajar en la intersección de la ciencia de datos y las ciencias sociales, colaborando estrechamente con profesionales de tecnología, ingenieros, y expertos en otros dominios específicos de manera colaborativa y no competitiva. La interdisciplinariedad es clave para abordar problemas complejos y globales: tener la capacidad de comunicar y producir a través de diferentes idiomas y conocimiento es una característica esencial para el ejercicio de la actividad.

Refiriendo al buen proceso de comunicación exigido, los analistas de datos deben ser capaces de explicar sus conclusiones, la validación de modelos y los resultados de manera comprensible para técnicos y no técnicos. La transparencia y la explicabilidad son aspectos críticos, especialmente en contextos de actividades gubernamentales y de servicios públicos. Es necesario que los analistas incorporen el imperativo de la cuantificación y visualización de datos para que la efectividad e implementación real de sus proyectos sean garantizadas.

En resumen, el analista socio-computacional de datos en la sociedad computacional se beneficia de herramientas tecnológicas contemporáneas, a la vez que se ve en el deber de abordar nuevas consideraciones éticas y sociales, colaborar interdisciplinariamente y comunicar resultados de manera más accesible y transparente. Su papel se ha expandido y especializado en respuesta a la complejidad de la era de la información y la tecnología. Este conjunto de actividades, meticulosamente coordinadas por el profesional del futuro en ciencias sociales, traza una ruta esencial hacia la construcción de un entorno socio-tecnológico equitativo y sostenible. Podemos concluir, que el rol del analista se presenta como el más complejo. A pesar de haber desempeñado este papel a lo largo de nuestras funciones tradicionales, la introducción de técnicas y del lenguaje computacional en el análisis de datos transforma significativamente la naturaleza de este trabajo.

El analista debe mantenerse constantemente actualizado en cuanto a las metodologías cualitativas, pero también debe adentrarse en la visualización de gráficos y en el vocabulario

específico de esta nueva era, entre otros aspectos previamente mencionados. No sólo cambian la forma y la cantidad de datos que se manejan, sino que la escala masiva añade complejidad al análisis, exigiendo una capacidad de adaptación, conexión y teorización sobre nuevos conocimientos antes no contemplados.

La capacidad de analizar todo y a todos en cualquier momento implica la necesidad de recortar, seleccionar e identificar lo verdaderamente relevante para incorporar a los proyectos. Especializarse en la concepción de estos criterios de elección, determinar las variables involucradas y colaborar en el proceso de calidad de los datos junto a los técnicos externos del proyecto se convierte en un papel esencial que los analistas socio-computacionales deben desempeñar. Desde el control de errores hasta la validación de resultados y soluciones, se establece el grado de especialización necesario para transitar entre saberes aparentemente no conectados hasta entonces.

En la era de la digitalización de la vida, a los profesionales del futuro les corresponde el poder de análisis para, junto con las máquinas, predecir el futuro. Es precisamente este factor, junto con los significativos cambios no sólo en la sociedad, sino también en todas las ciencias y ciencias sociales, lo que impulsa la emergencia de nuevos oficios y profesiones, generando la necesidad de introducir y respaldar académicamente estos nuevos saberes asociados a metodologías y paradigmas de investigación innovadores.

El surgimiento de nuevas profesiones y oficios está llevando la estructuración académica de estos saberes, evidenciado por la aparición de centros de investigación en sociología digital, programas de estudio y posgrados en ciencias sociales computacionales y humanidades digitales en todo el mundo, especialmente en América Latina. Además de confirmar la necesidad de mano de obra calificada, se refleja el reconocimiento institucional de la urgencia de incorporar estos conocimientos del futuro en la educación superior.

En el siguiente y último apartado, concluiremos el recorrido de este trabajo presentando una consecuencia del desarrollo de la sociedad computacional y de las ocupaciones sociales en torno a esta. Abordaré cómo el surgimiento de nuevas actividades y roles en la práctica de las ciencias sociales da origen a la necesidad de institucionalizar dichos conocimientos. En los últimos cinco a siete años, hemos observado la aparición de “disciplinas computacionales” en el ámbito de las humanidades. Finalmente, discutiremos cómo el surgimiento de nuevas formas de trabajar computacionalmente la información genera la necesidad de

institucionalizar el conocimiento y crear un sistema formal de educación que respalde, forme y estructure la profesión del futuro en las ciencias sociales.

### **10.3 La Creación académica de las Humanidades Digitales y el saber computacional**

En la contemporaneidad, la intersección entre sociedad y tecnología ha alcanzado parámetros inéditos, marcando el surgimiento de lo que he denominado de sociedad computacional. Este fenómeno, impulsado por la digitalización, la automatización y la omnipresencia de la tecnología computacional, ha redefinido radicalmente el tejido mismo de nuestras interacciones sociales.

En este contexto de transformación acelerada, la profesión del científico social emerge como un elemento crucial para descifrar los matices y complejidades de una sociedad que se desenvuelve en la confluencia de lo análogo y lo digital. En el análisis previo, se delineó con minuciosidad la figura del "científico social computacional", vislumbrando las habilidades y competencias que caracterizan a este profesional del mañana.

Este escenario plantea una pregunta ineludible: ¿Cómo preparar a los científicos sociales del presente para enfrentar los retos y aprovechar las oportunidades que ofrece la sociedad computacional? La respuesta puede residir en la creación de programas académicos especializados que aborden de manera integral las demandas emergentes del mercado laboral.

Con respecto a este punto, pasamos a observar un creciente proceso de institucionalización de dichos conocimientos, al interior de la academia, por parte de numerosos programas de ciencias sociales alrededor de latinoamérica y hasta mismo el mundo. Lo que no es raro de esperar. La propia Knorr Cetina (2005) discute brevemente este fenómeno a investigar cómo el conocimiento científico es "fabricado" y legitimado dentro de los laboratorios y cómo este proceso es moldeado por interacciones sociales y contextos específicos. Estos procesos revelan dinámicas importantes que pueden influir en la creación de nuevos campos o disciplinas. Por ejemplo, la manera en que ciertos conocimientos y métodos específicos se consolidan y se diseminan entre redes de científicos, así como la adaptación de tecnologías y prácticas, puede preparar el terreno para el surgimiento y la formalización de nuevos campos disciplinares.

En el mismo sentido, también podríamos traer como soporte teórico las discusiones de Latour (1987) en su libro *Ciencia en Acción*, donde defiende la institucionalización de una disciplina no solamente como un proceso de aceptación teórica, sino una negociación continua que implica crear alianzas, conquistar recursos y movilizar apoyo. Él argumenta que, para que una nueva disciplina emerja, los científicos deben construir y mantener una red robusta que dé legitimidad a su práctica. Aquí se enfatiza los cambios en la práctica laboral de los productores de conocimiento- exactamente lo que se evidencia en la discusión propuesta en este trabajo: cambios en la práctica profesional generando necesidades de consolidación del saber a partir de la institucionalización de nuevas disciplinas.

Adelante se habla sobre la emergencia y la creación de cursos, posgrados y diplomados denominados "Ciencias Sociales Digitales" o "Humanidades Digitales". De hecho, este movimiento ya ha empezado a despuntar en diferentes países de América Latina y en el mundo, principalmente a partir de los años 2010. Este llamado a la institucionalización como una disciplina surge como una respuesta en pos de la especialización del mercado de trabajo, donde la interconexión entre lo social y lo tecnológico demanda habilidades y conocimientos específicos.

La elección del término "humanidades digitales" en lugar de "ciencias sociales digitales" (más popular y frecuente en los programas de posgrado en América Latina) está arraigada en los orígenes y la propuesta multidisciplinaria de las humanidades digitales. Este campo integra enfoques y métodos digitales no solo para cuestiones tradicionales de las ciencias sociales, sino también para una gama de disciplinas centradas en el estudio de las expresiones culturales, históricas y literarias. Por ejemplo, áreas como la historia, la filosofía, la lingüística, la literatura y los estudios culturales han contribuido a la base de las humanidades digitales, ampliando su alcance e incorporando metodologías que tradicionalmente no se asociaban directamente con las ciencias sociales (Walsh, John *et al.* 2021)

Otro punto importante es el histórico institucional: las humanidades digitales comenzaron a consolidarse con base en las necesidades e intereses de las humanidades clásicas al adaptarse al entorno digital, explorando métodos como el análisis de textos, archivos digitales y la visualización de datos para estudiar y preservar el patrimonio cultural. Esto generó un contexto académico que buscaba preservar la identidad de las humanidades, incluso mientras se adaptan a nuevas tecnologías, por lo que se prefirió el término "humanidades digitales" para marcar este origen y diferenciación frente a las ciencias sociales (Grigar, Dene. 2013).

Además, el término "humanidades digitales" permite una flexibilidad que "ciencias sociales digitales" podría limitar, precisamente porque puede aplicarse a una gama más amplia de áreas y fomentar un espacio de experimentación metodológica e interdisciplinaria, algo esencial para una disciplina emergente (Kurniasih, Nia. 2021).

A continuación, se explicará las razones fundamentales que respaldan dicha adaptación curricular, examinando cómo las instituciones educativas pueden transformar sus programas para cultivar profesionales equipados no sólo con las herramientas tradicionales de la investigación social, sino también con las destrezas digitales y la agilidad necesaria para abrazar el futuro con confianza. Presentaré las principales características y formato curricular de estos programas, en base a la investigación documental que he realizado, principalmente a partir de 2022. El objetivo es describir el escenario actual de estas especialidades emergentes, demostrando las especificidades de su producción de conocimiento y justificativas para su difusión y enraizamiento en el campo de las humanidades contemporáneas.

### ***10.3.1 La computacionalización de los saberes sociales: ¿Qué son las Humanidades Digitales?***

Humanidades digitales, no es un concepto nuevo. Los primeros principios de la disciplina fueron difundidas por un pionero que, de alguna manera, sentó las bases de lo que hoy conocemos como las ciencias sociales digitales: el jesuita italiano Roberto Busa, quien durante la década del cuarenta del siglo XX, aprovechó las capacidades del ordenador para llevar a cabo la lematización de las obras de Santo Tomás de Aquino, estableciéndose así como el fundador de las humanidades digitales (Hockey, 2004). Ya desde entonces, el trabajo con corpus lingüísticos constituye una de las áreas más trabajadas en las humanidades digitales.

En una época marcada por la Segunda Guerra Mundial, entre 1941 y 1946, mi búsqueda incansable por máquinas que automatizaran el análisis lingüístico de textos escritos comenzó. Fue en 1949, en las instalaciones de IBM en la ciudad de Nueva York, donde encontré las herramientas que marcarían el inicio de una revolución. Como un anciano patriarca nacido en 1913, me asombro ante los desarrollos enormemente mayores y mejores de lo que en aquel entonces podría haber imaginado. ¡Digitus Dei est hic! ¡El dedo de Dios está aquí! (Winter,1999, p.6)

A principios del siglo XX, surgió una conexión entre el progreso humano y el avance tecnológico, lo que abrió la posibilidad de abordar interrogantes que se planteaban desde hacía mucho tiempo. Tras la Segunda Guerra Mundial, se iniciaron experimentos con el uso de tecnologías computacionales y el análisis de elementos culturales, especialmente en el ámbito literario. En la década de 1990, la revolución de las computadoras personales y la World Wide Web (WWW) abrieron las puertas para la configuración de lo que hoy conocemos como las humanidades digitales (Rojas Castro, 2013).

Un hito destacado en el surgimiento de las humanidades digitales fue la creación del consorcio Text Encoding Initiative en 1987, cuyas recomendaciones para codificar textos electrónicos vieron la luz en 1994. Durante los años noventa, la mayoría de los proyectos digitales se gestaron en Estados Unidos e Inglaterra (por ejemplo *Women Writers Project*, *Rossetti Archive* y *William Blake Project*), aunque en el ámbito hispanohablante también surgieron ejemplos tempranos como el ADMYTE<sup>116</sup>, por ejemplo (Rojas Castro, 2013).

Hasta mediados de los 2000, la disciplina se conocía como “humanidades computacionales”. El cambio de nombre, comúnmente atribuido a John Unsworth, editor del libro *A Companion to Digital Humanities* (2004), llevó a que en España, en los primeros años del siglo XXI, se usará la expresión “informática humanística” siguiendo el modelo italiano (Lucía Megías, 2003). A partir de 2003, en trabajos académicos en lengua española, empezamos a encontrar los primeros usos de la expresión “humanidades digitales”, la cual ha experimentado un cambio de nombre en diversas lenguas y regiones, facilitando su absorción masiva a partir de la generación de una conexión más estrecha entre la cultura humana y la tecnología (Saisó, 2008).

Hoy día, el concepto alrededor de la disciplina ya es un consenso establecido: las humanidades digitales es un área múltiple de investigación instrumental, aplicada y constituida por medio de su carácter digital en la que converge el uso de la informática con las ciencias humanas. Tienen como herramienta principal las nuevas tecnologías así como la utilización de plataformas digitales, desarrollos de Software y en general los medios digitales. Pueden organizar datos e información a gran escala, lo cual permite el acceso a una visión global del objeto de estudio.

---

<sup>116</sup>ADMYTE: Archivo Digital de Manuscritos y Textos Españoles. Base de datos conteniendo los textos medievales de España escritos en diversas lenguas peninsulares y más de 60 incunables e impresos que alberga la Biblioteca Nacional de textos antiguos españoles.

Representa, además, una reimaginación profunda de las disciplinas humanísticas, recobrando el propósito esencial de explorar el desarrollo humano en el contexto de su historia y cultura, aprovechando las nuevas posibilidades para la creación y difusión del conocimiento. Este enfoque no sólo revitaliza las humanidades, sino que también les devuelve su relevancia social.

La inclusión del término “digital” brinda una imagen menos técnica en comparación con el vocablo “informático”, pero también, por otro lado, esconde diversos atributos y características que van más allá de su rasgo de “ser digital”. Aun así, esta transición, según algunos autores como Spence (2014), es un camino conceptual acertado porque no solamente democratiza el acceso a conceptos técnicos en la sociedad, sino que también, al menos en castellano, subraya aún más la importancia de “las humanidades”, elevándose de adjetivo a sustantivo: las humanidades son el cimiento, la base, el digital, su adjetivación.

Las humanidades digitales abarcan una amplia gama de objetos de estudio, desde el diseño y mantenimiento de colecciones digitales hasta el análisis a gran escala de datos culturales. Comúnmente referidas con la abreviatura HD, esta disciplina contemporánea engloba tanto materiales y patrimonio analógico convertido al digital como artefactos ya nacidos en el entorno de la digitalidad. En su convivencia, se apropia de las metodologías inherentes a las disciplinas humanísticas tradicionales (historia, filosofía, filología, lingüística teórica, lingüística aplicada, historia del arte, arqueología, música y estudios culturales) y las ciencias sociales, con el uso de herramientas informáticas (bases de datos, visualización de datos, recuperación de información, minería de datos, estadística) y la edición digital de todas estas actividades (Hockey, 2004).

Cuando aplicadas a la investigación, las humanidades digitales se entrelazan estrechamente con el concepto de *e-Research*, estableciendo relaciones de aplicación y epistemológicas que modelan la interpretación de la realidad. Esto se manifiesta a través de bases de datos, gráficos, mapas y ediciones electrónicas, entre otras formas (Hockey, 2004). En este renacer de las disciplinas humanísticas, se presentan como el puente que conecta el pasado, el presente y el futuro, revitalizando la comprensión de la condición humana en la era digital (Berry, 2011). Más allá de incorporar nuevas herramientas a la esfera de las humanidades, se establece un diálogo interdisciplinario que da lugar a un movimiento cohesivo y receptivo. Según David Berry (2011), producimos un *Giro Computacional* en las humanidades cuando las técnicas computacionales influyen en todos los aspectos de estas disciplinas. Estas

técnicas no actúan simplemente como instrumentos complementarios a los métodos tradicionales, sino que afectan a todos los aspectos de las disciplinas, introduciendo nuevos enfoques para identificar patrones en los datos más allá de las narrativas y comprensiones convencionales. Esto, a su vez, posibilita la modularización y recombinação de las disciplinas, trascendiendo el entorno académico tradicional. Esta comunidad académica tiene como objetivo central revitalizar la epistemología y los métodos analíticos, haciendo que la investigación sea compatible con las necesidades sociales.

Este es justamente uno de los objetivos específicos de las humanidades digitales: la disciplina aspira a llevar a cabo la transformación de elementos tangiblemente humanísticos al ámbito digital, con el propósito de preservar y difundir este conocimiento. También se centran en la creación de bases de datos a partir de recursos digitales. Estas bases de datos buscan desarrollar nuevas estrategias de trabajo que faciliten la creación, organización y análisis de información, con el fin de generar conocimiento que promueva el avance en el estudio de las humanidades y su adecuación a los tiempos contemporáneos.

### ***10.3.2 Los currículum: la estructuración académica de las humanidades digitales***

No Brasil, o termo Humanidades Digitais ganhou adesão a partir da década de 2010 e passou a figurar, cada vez mais frequentemente, em publicações, eventos, nomes de cursos e de associações. A pesquisa visa discutir a experiência brasileira de inovação do ensino de nível superior a partir da implementação do enquadramento teórico-metodológico das Humanidades Digitais. Que tipos de formação têm sido privilegiados e em que áreas? Alguns dos aspectos cruciais da prática das Humanidades Digitais, do local ao global, incluem a mudança na natureza do treinamento dos alunos de graduação e de pós-graduação da grande área das Humanidades, os tipos de parceria e de pedagogia interdisciplinar que passam a ser mais comuns e desejadas, o incentivo a práticas colaborativas e equipes interdisciplinares, e a orientação da pesquisa ao público mais amplo, para além dos muros das universidades. Ao entendermos as transformações que têm ocorrido no campo das Humanidades no contexto nacional, em resposta à crescente digitalização do mundo cotidiano e da nossa prática profissional, estaremos melhor preparados para atuar na construção de caminhos que fortaleçam boas práticas científicas, a cidadania e a valorização do conhecimento

humanístico e crítico”<sup>117</sup>. (LHuD<sup>118</sup>, 2023 Página web: Laboratório de Humanidades Digitais da FGV CPDOC - accedido en octubre 2023)

Esta cita expresa el cerne de lo que me interesa presentar en esta parte: la actual relevancia de la institución de experiencias que formalizan y consolidan el tipo de conocimiento que estamos fabricando, y en consecuencia, de nuestro existir en la sociedad computacional. La investigación se enfoca en la experiencia brasileña de innovación en la educación superior a través de la implementación de las humanidades digitales. Nos cuenta sobre una promesa, de una profesión favorecida, innovadora y valorada, por destacar cambios cruciales en las áreas con las cuales trabaja e impacta. Sobre todo en la educación de estudiantes de grado y posgrado en humanidades y a partir de pilares como la interdisciplina y las prácticas colaborativas.

El texto que materializa la presentación pública de los objetivos de la investigación en curso, en la Fundación Getulio Vargas, una de las instituciones de posgrado más renomadas de Brasil y Latinoamérica, expresa de forma evidente algunos de los principios de incidencia, colaboración, identidad global, interdisciplinar, y por supuesto, el tecno encanto, presentada por este proyecto. Como ya sabemos, esta misma temática se encuentra bastante presente no sólo en la literatura de referencia que he presentado en este trabajo, sino también en diversas investigaciones que ocurren en este exacto momento: versiones diferentes del mismo fenómeno. Les acabo de presentar una investigación en curso en Brasil, también está la mía, las investigaciones de Henry, mencionado compañero que encontré en el camino y toda la gente por detrás de su laboratorio en Ecuador. Está el propio laboratorio que estudié y del que soy parte, GobLab, y así decenas de laboratorios, núcleos e investigadores sociales cada vez más interesados en la implementación de técnicas prácticas y teorías digitales y computacionales en latinoamérica.

“Mapear el encino que se auto identifique en el dominio das Humanidades Digitales o que aplique métodos digitales y abordagens computacionales a cuestiones, desafíos, datos y fuentes de pesquisa propias de las Ciencias Humanas” (Laboratório de Humanidades Digitais

---

<sup>117</sup>La citada investigación forma parte del proyecto "Alfabetización digital: modelización de competencias digitales para humanistas y científicos sociales", financiado por el Fondo de Investigación Aplicada de la Fundación Getulio Vargas y coordinado por Jimmy Medeiros, Suemi Higuchi y Juliana Marques, investigadores de la Facultad de Ciencias Sociales del CPDOC de la FGV.

<sup>118</sup><https://cpdoc.fgv.br/laboratorios/lhud>

(LHuD), 2023). Esto resume, en gran parte, la propuesta de esta sección, que es una recolección de currículum de diferentes posgrados con los que he tenido contacto a lo largo de esta investigación.

En la ocasión de una de mis visitas a Brasil, conocí el Programa de Posgrado Interdisciplinar de la Universidad Federal Rural de Rio de Janeiro (UFRRJ), responsables por el magíster en Humanidades Digitales, pionero en el país. Tuve la oportunidad de conversar con algunas personas, estudiantes y cuerpo docente, durante un congreso internacional de humanidades digitales, organizado por el mismo programa. Luego de estas conversaciones y lectura de sus mallas curriculares, detecté un patrón que resumo a continuación.

Por lo general los programas de posgrado son de tipo magíster, con la duración de 2 años, 4 semestres. Son 60 créditos de introducción a las humanidades digitales y 30 créditos de disciplinas más técnicas o metodológicas. Más 30 créditos de disciplinas optativas variadas. Normalmente estos dos años se dividen entre las disciplinas obligatorias: Introducción a las Humanidades Digitales, Métodos Computacionales en Humanidades Digitales, y los seminarios en Humanidades Digitales, que corresponden a créditos en temáticas en específicas ofertadas por los programas.

En las disciplinas de introducción a las humanidades digitales suelen haber varias combinaciones de seminarios según la oferta del programa y del interés de los estudiantes. En general, este módulo introduce las humanidades digitales y examina cómo aportan teoría y práctica digital al estudio de la cultura humana, desde la historia, las lenguas y la música hasta los museos, la edición digital y más allá. Consiste en una formación básica en los conceptos que definen el campo de estudio de las humanidades digitales, dirigido tanto a estudiantes de Informática como de Humanidades o Ciencias Sociales. En estos tipos de cursos, se abordan e identifican los aspectos que definen los nuevos horizontes en el estudio de las Humanidades y las Ciencias Sociales en la era de los registros digitales masivos del comportamiento humano.

Los temas tratados se clasifican y presentan con un enfoque crítico, por lo general, destacando la importancia de combinar análisis cualitativo y cuantitativo. Se presta especial atención a la aplicación del conjunto de métodos computacionales de éxito e inaccesibles antes de la era digital: la representación de datos complejos, la visualización de la información y la construcción y estimación de modelos. En este contexto, se reexamina el

método científico en humanidades o ciencias sociales, explorando cómo la informática y los datos digitales permiten nuevas respuestas a investigaciones clásicas, el planteamiento de nuevas preguntas y nuevos retos y oportunidades éticos.

En el programa de postgrado en humanidades digitales de la UFRRJ, entre los cursos introductorios también se encuentra la discusión y aprendizaje sobre los conjuntos de métodos y prácticas computacionales, donde ven cómo el conjunto de oportunidades y retos de la integración de la informática y las humanidades, pueden beneficiar la producción de conocimiento social en general. Posee, dentro de este ramo básicamente tres fuertes líneas de foco:

1- Datos sociales en formato digital: observación de comportamientos sociales, recopilación de datos sociales, rastreo de documentos, organización digital de datos sociales, estrategias de investigación basadas en grandes cantidades de datos digitales.

2- Análisis automático de conjuntos de datos: principios del análisis de contenido, semántica frente a sintaxis, posibilidades de aplicación de técnicas de minería de datos, formulación de preguntas para investigar sobre un conjunto de datos, categorías de análisis, estudios de casos.

3- Simulación de la dinámica social: conceptualización de la investigación de modelos sociales complejos mediante modelos, sistemas sociales dinámicos, modelización formal, comparación de casos de aplicación de modelos de autómatas celulares, estocásticos o basados en agentes, estudios de casos. Práctica fundamental, como estudios observacionales y experimentos, resúmenes estadísticos, visualización de datos y cómo las oportunidades computacionales pueden mejorarlos.

La segunda disciplina obligatoria (Métodos Computacionales en Humanidades Digitales) consiste en un enfoque más técnico y metodológico a los desafíos del trabajo con los datos e información sociales analizados de forma computacional. Enseñará a los estudiantes las estructuras básicas y la sintaxis de un lenguaje de programación común. Comprende una amplia descripción de los diversos métodos computacionales aplicados en análisis cuantitativos de datos sobre fenómenos o comportamientos sociales. La presentación y discusión de los conceptos de estos métodos computacionales pretende construir la visión crítica necesaria para poder adaptarlos a casos concretos desde las humanidades o las ciencias sociales.

Los métodos son presentados en términos de sus principios, definiciones y contextos de aplicación, formando una visión general de los mismos. Siguiendo este enfoque, este curso prepara a los estudiantes con formación en informática, en humanidades o ciencias sociales para formular problemas computacionales que modelen un problema concreto, identificar un método computacional adecuado para resolverlo y calibrar este método. Aunque por lo general este tipo de disciplina no se trata de una formación de programadores especializados, sí aborda el conocimiento de cómo adaptar y ampliar las habilidades computacionales a cuestiones relacionadas a las humanidades y ciencias sociales, a medida que se presentan nuevas preguntas, métodos y datos. Los conceptos claves abordados en el transcurso de las clases son: ordenadores, algoritmos y programas, operadores y control de flujo, estructuras de datos, análisis de la complejidad, lenguajes de programación, compiladores e intérpretes, orientación a objetos, UML. Los principales temas discutidos al interior de la disciplina son:

1- Métodos computacionales para la minería de datos: problemas de agrupación y clasificación, análisis de frecuencia de ocurrencias y reglas de correlación, análisis semántico y de sentimiento de textos, preprocesamiento de datos, modelado de grafos, calibración, estudios de casos en laboratorios informáticos utilizando software existente.

2- Métodos computacionales para la simulación: fundamentos de la complejidad, autoorganización, emergencia, imprevisibilidad, simulación; definición básica de autómatas celulares; adaptaciones de los autómatas celulares a los problemas derivados de la dinámica social; definición de modelos multiagente; adaptaciones de los modelos multiagente a los problemas derivados de la dinámica social; estudios de casos en laboratorios informáticos utilizando software existente para autómatas celulares y modelos multiagente.

Por lo general este es un ramo práctico y aunque los estudiantes no tengan experiencia previa en codificación o programación, adquieren bagaje conceptual y técnicas básicas que podrán aplicar a una variedad de materiales relacionados con el estudio de la cultura humana. A través de talleres prácticos y ejercicios de diseño de proyectos en ciencia de datos, y de tecnologías computacionales, aprenderán a crear proyectos de dicha naturaleza. Estas competencias son transferibles y suficientes para poder utilizar más allá del ámbito académico.

Los diplomados del Laboratorio de Innovación Pública de la UAI, GobLab, son según las formaciones mapeadas, es la única iniciativa entre las que investigué que efectivamente enseña conocimientos básicos de programación a todo alumno.

Llegan a nosotros profesionales de empresas que no tienen conocimientos en programación. Nuestro diplomado se centra en enseñar ciencia y datos sin requerir conocimientos previos de programación. Esto satisface una demanda, ya que las escuelas de ingeniería suelen asumir que los estudiantes deben aprender a programar por sí mismos. Nuestro enfoque es desde cero; sin embargo, a pesar de ser accesible, el diplomado es 100% técnico, con todas las evaluaciones realizadas en R, un lenguaje ampliamente utilizado en ciencias sociales debido a su amigabilidad. Aunque actualmente nos centramos en R, en proyectos de investigación aplicada utilizamos Python, ya que algunos profesores están familiarizados con ambos. Creemos que en el futuro, Python podría reemplazar a R, pero enseñar dos lenguajes a profesionales de ciencias sociales no es práctico. Los informáticos pueden adaptarse fácilmente a varios lenguajes, pero para aquellos no familiarizados con la programación, aprender un solo lenguaje ya es un desafío considerable. (MPH - Entrevista 32/2022 - Chile. Mi traducción)

Como derivación de este relato se puede identificar la existencia de una real demanda para aprender ciencia de datos, especialmente para aquellos profesionales que no tienen conocimiento previo de programación. La falta de opciones que aborden esta necesidad específica lleva a que el diplomado mencionado, que inicia desde cero, sea popular, pues permite que incluso aquellos sin experiencia previa en programación adquieran habilidades en ciencia de datos.

Se ve, además, una nueva tendencia en enseñar R en Ciencias Sociales, y aunque la formación actual se centra actualmente en R debido a su accesibilidad y amigabilidad para las ciencias sociales, se reconoce que muchos proyectos de investigación aplicada utilizan Python, por lo que hay un creciente interés por aprender esta segunda herramienta. Se prevé que la enseñanza de Python se integre gradualmente en el futuro para mantenerse al día con las tendencias y demandas del sector.

Es importante mencionar que se habla de los desafíos de enseñar múltiples lenguajes a profesionales no informáticos como a los de ciencias sociales, quienes muchas veces perciben la tarea de aprender a programar como “abrumadora”. El cambio entre lenguajes es más fácil para los informáticos, pero para quienes no tienen experiencia en programación, la transición

a otro lenguaje requiere un intenso proceso de estudio, como lo evidencian los estudiantes que, después de un tiempo trabajando en R, deciden aprender Python. La elección de un lenguaje se basa en la practicidad y la comodidad de los estudiantes, y se especula que, con el tiempo, Python podría reemplazar a R.

El tercer y último ramo obligatorio, los seminarios en Humanidades Digitales, consisten en espacios de enseñanza destinados a fomentar la interdisciplinariedad y la formación como investigador en temas actuales que integran la informática y las humanidades. Los alumnos deben presentar una solución relativa a los temas centrales de investigación siendo conducidos a debatir cuestiones y temas de actualidad que integran ambas disciplinas con investigadores u otro tipo de invitados. Esta instancia busca cautivar a los asistentes a vincularse en una temática o vivencia investigativa específica. También presentan desarrollos relativos a sus temas centrales de investigación en forma de seminarios.

No todos los programas poseen la tesis con determinación para recibir el título de grado. En más de tres casos, vi que un determinado número de publicaciones en revistas indexadas sustituye el requisito de la tesis. En otro dos casos, para concluir el grado, aparte de cumplir con los créditos mínimos exigidos, el estudiante debe ser incorporado en algún proyecto de investigación aplicada, trabajando para el tercer sector, sea a través de una consultora, o una relación laboral formal.

Con respecto a las disciplinas optativas, como se puede imaginar, la variedad de ellas es aún más grande. Abajo presento una descripción general con algunas de las disciplinas que encontré durante mi investigación:

**Tabla 42.** *Listado de disciplinas optativas en posgrados en Humanidades Digitales.*

Disciplinas en cursos optativos más frecuentes	Descripción
<b>Humanidades Digitales e Investigación Geográfica</b>	Introduce a los estudiantes en la investigación geográfica, aplicando métodos computacionales. Analiza la geografía de la innovación, los datos geográficos, la cartografía, los sistemas de información geográfica, las geografías alternativas y las posibilidades digitales.
<b>Análisis y Extracción/minería de datos</b>	Explora los conceptos matemáticos y computacionales de las técnicas de minería de datos, centrándose en grandes volúmenes de datos. Incluye temas como la recopilación e ingeniería de datos, técnicas de minería de datos (clasificación, agrupación en clústeres, reglas de asociación), implementación práctica en R y casos prácticos (retención de clientes, detección de fraudes, sistemas de recomendación, análisis de redes sociales).

<b>Modelización y simulación de fenómenos urbanos</b>	Abarca técnicas de modelización y simulación para estudiar el desarrollo urbano, con especial atención a los autómatas celulares y los modelos basados en agentes. Incluye aspectos teóricos de la modelización urbana, su aplicación a las políticas públicas y su implementación práctica.
<b>Análisis y diseño de algoritmos</b>	Centrado en la formación en algoritmos, abarca desde modelos informáticos hasta técnicas y algoritmos polinómicos (recursividad, divide y vencerás, programación dinámica, algoritmos codiciosos), así como estructuras de datos básicas y teoría de la complejidad.
<b>Introducción al análisis geoespacial</b>	Explora el análisis de datos geoespaciales utilizando herramientas de software geoespacial de código abierto. Incluye la implementación de estructuras de datos espaciales, geovisualización, análisis de autocorrelación espacial, variografía, detección de conglomerados y minería de datos espaciales.
<b>Fundamentos de la ciencia de datos</b>	Abarca conceptos matemáticos como vectores, matrices, sistemas de ecuaciones lineales, determinantes, espacios vectoriales, transformaciones lineales, valores y vectores propios, factorización de matrices, formas lineales y cuadráticas, matriz definida positiva y resolución numérica de sistemas de ecuaciones lineales.
<b>Bancos de Datos</b>	Explora el diseño y desarrollo de bases de datos relacionales, SQL, formas normales, álgebra relacional, lenguaje de consulta SQL, estructuras de indexación, conectividad web, Big Data y operaciones MapReduce, NoSQL y NewSQL.
<b>Introducción a la Historia Digital</b>	Presenta la Historia Digital como un campo emergente en la disciplina histórica. Incluye el uso de herramientas digitales para preservar, analizar e interpretar la historia, con un enfoque en el pensamiento y la participación con herramientas digitales para producir expresiones digitales de los conocimientos adquiridos.
<b>Análisis y minería de medios sociales</b>	Abarca aspectos teóricos y metodológicos de la recogida, visualización, análisis e interpretación de datos en redes sociales. Incluye fundamentos matemáticos, teoría de grafos, álgebra lineal, estadística, análisis de comunidades, difusión de información, modelización y medición de redes y temas de visualización de datos.
<b>Tecnologías digitales inclusivas</b>	Analiza el desarrollo de tecnologías educativas digitales inclusivas, combinando conceptos computacionales con aspectos pedagógicos. Aborda temas como el pensamiento computacional, la alfabetización digital, la programación visual, las estrategias inclusivas, el diseño universal en el aprendizaje y la creación de libros digitales accesibles.
<b>Tecnología de la información jurídica</b>	Presenta las posibilidades de aplicación de las nuevas tecnologías de la información al tratamiento y análisis de datos jurídicos. Incluye temas como Derecho y Nuevas Tecnologías de la Información, Big Data aplicado al Derecho, sistemas de apoyo a la toma de decisiones jurídicas y gobierno electrónico en la justicia brasileña.
<b>Procesamiento del lenguaje natural</b>	Explora la manipulación computacional y cuantitativa de textos, incluida la limpieza, la anotación lingüística automática y métodos cuantitativos como la visualización de frecuencias y el modelado de temas. Abarca conceptos de procesamiento del lenguaje natural, correctores de escritura, clasificación de textos, análisis de sentimientos, identificación de entidades con nombre, extracción de relaciones y aplicaciones en Humanidades Digitales.

Esta tabla revela la amplitud y diversidad de disciplinas que se pueden cursar en el ámbito de las humanidades digitales. La interdisciplinariedad es una de ellas ya que abarcan una amplia gama de disciplinas, desde ciencias de la computación y estadísticas hasta geografía, historia,

educación, derecho y más. Este enfoque destaca la naturaleza colaborativa y la convergencia de diferentes campos para abordar problemas y preguntas complejas.

Otro factor a destacar es la variedad de temas discutidos y trabajados: incluyen análisis y minería de datos, análisis de medios sociales, diseño de algoritmos, bases de datos, ciencia de datos, investigaciones geográficas, análisis geoespacial, historia digital, modelado urbano, procesamiento de lenguaje natural, tecnología de la información jurídica y tecnologías digitales inclusivas. Esto destaca la versatilidad de las humanidades digitales y su capacidad para aplicarse en una amplia variedad de contextos, que además están alineados a desafíos y temáticas muy contemporáneas.

Las disciplinas abordadas en este programa de magíster en Humanidades Digitales están diseñadas para enfrentar desafíos contemporáneos, como el análisis de grandes conjuntos de datos, la comprensión de fenómenos urbanos, la preservación y análisis de información histórica digital, y la aplicación de tecnologías digitales inclusivas en la educación, a través de un cuerpo docente también diverso. Esto subraya la riqueza de perspectivas que los estudiantes pueden encontrar, así como las múltiples habilidades y conocimientos que pueden adquirir a lo largo de su formación.

En relación con la formación del cuerpo docente, se destaca una notable variabilidad y un alto grado de interdisciplinariedad. Dada la diversidad de disciplinas y enfoques abordados en los programas de posgrado, resulta imperativo que el cuerpo docente refleje esta diversidad. La consecución de una estructura de enseñanza interdisciplinaria y diversificada, que atienda los variados intereses de los alumnos, demanda que los profesores también tengan distintos perfiles académicos.

Pese a lo anterior, se identifica una marcada preferencia por ciertos temas y áreas a ser trabajadas. La figura 56 a continuación, consiste en una nube de palabras, resultado de un análisis sobre la formación y especialización de los académicos, extraído de los currículos de posgrado que he tenido la oportunidad de estudiar.

Como se puede apreciar, algunos términos son más frecuentes que otros. En el núcleo de las especialidades destaca el vínculo con los algoritmos, la computación y los datos. Términos como "derecho", "aprendizaje", "educación", "geografía" y "grafos" reflejan áreas temáticas que tienden a estar más consolidadas en estos programas. Si observamos con detenimiento estas predominancias en dichas temáticas, podemos notar una conexión directa con los datos

presentados al inicio de este capítulo, donde se presentaron las especialidades y temáticas más comunes trabajadas por los científicos sociales contemporáneos involucrados en sistemas de investigación en ciencia de datos

**Figura 56.** Nube de palabras especialidad cuerpo docente



### 10.3.3 Áreas de aplicación: usando las tecnologías computacionales para investigar la sociedad digital

Se enfocan en aprovechar las tecnologías digitales para enriquecer disciplinas humanísticas y culturales. En el ámbito del análisis de texto y corpus lingüísticos, por ejemplo, se emplean herramientas computacionales para explorar patrones lingüísticos y la evolución del lenguaje. También es posible analizar obras antiguas o incompletas a causa del deterioro, pueden ser digitalizadas y analizadas automáticamente, lo cual suele ser útil en historia, arte y cultura. En estas disciplinas, hay interés por la preservación y análisis de obras de arte digital o la digitalización de archivos históricos, por ejemplo.

Las geo humanidades, por otro lado, integran herramientas digitales en el estudio geográfico de fenómenos culturales. Los estudios literarios digitales implementan análisis computacional a textos literarios. La educación digital integra tecnologías digitales en la enseñanza, mientras que los estudios de medios digitales investigan fenómenos mediáticos y culturales en el entorno digital, para dar más ejemplos que ilustran la versatilidad de las humanidades digitales, que al unir tecnología y humanidades, enriquecen la comprensión y preservación de

la cultura y la historia. A continuación, una síntesis de algunas de las aplicaciones de los métodos y saberes computacionales en las humanidades digitales.

**Tabla 43.** *Aplicabilidades de los recursos computacionales en diferentes disciplinas de las humanidades digitales*

<b>Áreas donde implementar las humanidades digitales</b>	<b>Descripción</b>
<b>Análisis de Texto y Corpus Lingüísticos</b>	Utilización de herramientas computacionales para analizar grandes cantidades de texto, identificar patrones lingüísticos y explorar la evolución del lenguaje a lo largo del tiempo.
<b>Historia Digital</b>	Digitalización de archivos históricos, creación de bases de datos históricas y uso de herramientas digitales para analizar y visualizar eventos y tendencias históricas.
<b>Arte y Cultura Digital</b>	Preservación y análisis de obras de arte digital, así como la creación de experiencias interactivas para la apreciación y comprensión de manifestaciones culturales.
<b>Geo Humanidades</b>	Integración de herramientas digitales en el estudio geográfico y espacial de fenómenos culturales y humanos, como la cartografía digital y la visualización geoespacial.
<b>Estudios Literarios Digitales</b>	Análisis computacional de textos literarios, exploración de patrones narrativos y creación de ediciones digitales de obras literarias.
<b>Filosofía Digital</b>	Exploración de cuestiones filosóficas mediante herramientas digitales, como la simulación y modelado computacional para abordar problemas abstractos.
<b>Educación Digital</b>	Integración de tecnologías digitales en la enseñanza de las humanidades, incluyendo la creación de recursos educativos digitales y plataformas de aprendizaje en línea.
<b>Patrimonio Digital y Museos Virtuales</b>	Digitalización y preservación de patrimonio cultural, así como la creación de museos virtuales y experiencias inmersivas para acceder a la historia y la cultura.
<b>Estudios de Medios Digitales</b>	Investigación de fenómenos mediáticos y culturales en el entorno digital, incluyendo análisis de redes sociales, memes y otros aspectos de la cultura digital.
<b>Ciencias Sociales Computacionales</b>	Aplicación de métodos digitales para el análisis de datos sociales, incluyendo encuestas en línea, minería de datos y modelado computacional en estudios sociológicos.

Además de presentar ejemplos de las aplicaciones del saber computacional y metodologías digitales en los estudios, investigaciones y soluciones a problemáticas sociales, se señala la manera en que las humanidades digitales han revolucionado las ciencias sociales al introducir recursos computacionales en disciplinas como la sociología, la ciencia política y la antropología, historia, geografía, lingüística, comunicación, derecho, entre muchas otras.

La capacidad de analizar grandes conjuntos de datos mediante técnicas de big data y ciencias de datos, permite a los científicos sociales identificar patrones, relaciones y tendencias complejas que podrían pasar desapercibidos a través de los métodos tradicionales. Además, el modelado computacional y las simulaciones ofrecen una comprensión más profunda de fenómenos sociales complejos, como dinámicas de grupo y cambios en la estructura social. El procesamiento de lenguaje natural (NLP), por ejemplo, facilita el análisis de enormes cantidades de texto, revelando patrones lingüísticos, sentimientos y temas emergentes en discursos y redes sociales. Asimismo, las herramientas de visualización de datos, como gráficos interactivos y mapas, permiten presentar resultados de manera accesible, contribuyendo enormemente para la masificación de los resultados y el intercambio de información entre proyectos sean ellos científicos o no. Otro ejemplo es el machine learning y el aprendizaje automático posibilitan la predicción de comportamientos sociales y la mejora de la calidad de las decisiones, aunque requieren una consideración ética fundamental, abordando cuestiones de privacidad y sesgo algorítmico.

En este contexto, la colaboración interdisciplinaria entre expertos en ciencias sociales, informática y estadística se vuelve esencial. La profesionalización de los científicos sociales computacionales destaca la importancia de combinar habilidades de diversas disciplinas para abordar integralmente cuestiones complejas. En este campo en constante evolución, es imperativo enfatizar la ética y la responsabilidad en la aplicación de estos recursos computacionales para garantizar un impacto positivo y sostenible en la investigación social.

En este contexto, la profesionalización de los conocimientos computacionales en las ciencias sociales no recae únicamente en la academia, sino que se ve respaldada por la formalización de estos conocimientos a través de cursos, diplomados e iniciativas de posgrado. La academia desempeña un papel crucial en este proceso, pero también emergen iniciativas "formales" fuera de este ámbito, dando paso a la introducción de las humanidades digitales en la producción de conocimiento social. Observamos el surgimiento de organizaciones,

asociaciones e instituciones sin fines de lucro que, de manera paralela, contribuyen significativamente a este proceso.

A lo largo de mi investigación, he sido testigo del nacimiento de diversas redes de actores interesados en el estudio y profundización de la discusión digital y computacional dentro de sus disciplinas.

#### ***10.3.4 Las humanidades digitales en América Latina***

Finalmente, para facilitar la exploración de redes y formas de asociación en este ámbito de investigación, presento un listado de instituciones en humanidades digitales a nivel mundial y en América Latina que captaron mi atención durante la investigación. Es crucial destacar que esta lista representa sólo una pequeña muestra, resultado de búsquedas realizadas desde un computador registrado con una dirección IP específica, lo cual puede influir en la búsqueda, consecuencia, justamente, de la reducción de la espontaneidad de la vida. Esto es, los resultados arrojados corresponden a aquellos filtrados por los motores de búsqueda, y luego recalibrados por los algoritmos de los diferentes dispositivos y aplicaciones utilizadas para llevar a cabo esta investigación.

A pesar de esto, es una selección significativa influenciada por mi perspectiva y las decisiones conjuntas entre mi computadora, la internet y los algoritmos que respaldan estos procesos. Aunque limitado, este listado sigue siendo valioso como un registro y una suerte de arqueología de un contexto en el que emergen disciplinas, programas, redes y asociaciones, que sin duda se han multiplicado, ya que reconocemos que la velocidad es un principio clave en la sociedad computacional y está arraigada en los cimientos del desarrollo contemporáneo de la investigación social.

**Tabla 44.** *Listado de organizaciones en humanidades digitales*

<b>País</b>	<b>Organización</b>	<b>Año de fundación</b>	<b>Enlace de acceso</b>
Internacional	Alliance of Digital Humanities Organizations (ADHO)	2013	<a href="https://adho.org/">https://adho.org/</a>

+ 100 centros situados en 19 países	<i>CenterNet</i>	2009	<a href="https://dhcenternet.org/">https://dhcenternet.org/</a>
México	Red de Humanidades Digitales (RedHD)	2011	<a href="http://humanidadesdigitales.net/">http://humanidadesdigitales.net/</a>
España	Asociación Humanidades Digitales Hispánicas (HDH)	2011	<a href="https://humanidadesdigitaleshispanicas.es/">https://humanidadesdigitaleshispanicas.es/</a>
Argentina	Asociación Argentina de Humanidades Digitales (AAHD)	2013	<a href="https://aahd.net.ar/">https://aahd.net.ar/</a>
Argentina	Instituto de Humanidades Digitales de la Universidad Nacional de La Plata	2015	<a href="https://idihcs.fahce.unlp.edu.ar/">https://idihcs.fahce.unlp.edu.ar/</a>
República Dominicana	Observatorio de las Humanidades Digitales, UASD	2015	<a href="https://www.facebook.com/Humanidadesdigitales/?locale=es_LA">https://www.facebook.com/Humanidadesdigitales/?locale=es_LA</a>
Puerto Rico	Observatorio de Humanidades Digitales	2016	<a href="https://cea.uprrp.edu/ciclo-de-las-humanidades-digitales/">https://cea.uprrp.edu/ciclo-de-las-humanidades-digitales/</a>
Colombia	Red Colombiana de Humanidades Digitales	2016	<a href="https://rhd.com.co/">https://rhd.com.co/</a>
Colombia	Investigación de Humanidades Digitales de la Biblioteca Nacional de Colombia "De Punto a Pixel"	2017	<a href="https://bibliotecanacional.gov.co/es-co/servicios/en-la-bnc/grupo-de-investigacion/proyecto-punto-pixel">https://bibliotecanacional.gov.co/es-co/servicios/en-la-bnc/grupo-de-investigacion/proyecto-punto-pixel</a>
Brasil	Grupo de investigación en Humanidades Digitais	2011	<a href="https://humanidadesdigitais.org/">https://humanidadesdigitais.org/</a>
Brasil	Associação das Humanidades Digitais	2013	<a href="http://abhd.org.br/">http://abhd.org.br/</a>
Venezuela	Red Historia Venezuela	2021	<a href="https://redhistoriave.org/">https://redhistoriave.org/</a>
Haiti	H-Haiti	2016	<a href="https://networks.h-net.org/h-haiti">https://networks.h-net.org/h-haiti</a>
Chile	Humanidades Digitales Chile	2019	<a href="https://www.facebook.com/HDigitalesChile/">https://www.facebook.com/HDigitalesChile/</a>

Para concluir, decir que hemos visto despuntar muchas organizaciones en América Latina en los últimos 10 años. Esta contribución en América Latina tiene la potencialidad de convertirse en un campo disciplinar similar al europeo o norteamericano, y hemos experimentado un relativo auge en los últimos años, principalmente después de la pandemia COVID-19.

Tal como viene pasando con otros campos científicos, sobre todo el matemático, económico, biológico y las ingenierías, las ciencias sociales están cada vez más abiertas a la exploración y al descubrimiento de las posibilidades que se instauran al quehacer científico a partir de utilización de las ciencias de datos, las tecnologías en datos masivos y automatizaciones (Nascimento, 2106). Hay una necesidad de acercarse a los procedimientos que envuelven estas ciencias con la finalidad de entender sus especificidades y potencialidades para entonces, ajustar, modernizar o complementar su mirada hacia al objeto de investigación por excelencia de las ciencias sociales: la sociedad.

Además, hemos visto inaugurar un nuevo subcampo disciplinario en las ciencias sociales, variando entre los nombres de sociología y antropología digital, las ciencias sociales computacionales o las humanidades digitales. Estas disciplinas han consolidado su presencia desde la primera década de los años 2010 y ofrecen formas innovadoras de abordar la investigación social.

Espero que este resumen ayude a ilustrar cómo las humanidades digitales se centran en la investigación de dinámicas sociales, desde individuos hasta grupos numerosos, mediante enfoques que combinan análisis cualitativos y cuantitativos. Su carácter innovador implica la manipulación de grandes cantidades o complejidades de información, utilizando conceptos, teorías y artefactos computacionales avanzados. Este campo aborda cuestiones en la intersección entre las ciencias humanas y de computación, ampliando el conocimiento teórico y aplicado de los métodos computacionales para comprender las interacciones humanas. Es una disciplina interdisciplinaria que aborda las complejidades de la Sociedad de la Información.

La constitución de las humanidades digitales responde a las transformaciones causadas por la difusión de tecnologías de información electrónica, afectando la generación, preservación y difusión del conocimiento sobre las sociedades humanas. También refleja el impacto de estas tecnologías en la labor de profesionales que enfrentan diversos problemas sociales, y la necesidad de reflexión crítica sobre los riesgos y desafíos de la omnipresencia de la tecnología digital en la vida humana.

### **Cap 11: Por unas ciencias sociales computacionales: conclusiones, aprendizajes y límites de la investigación**

El término big data fue acuñado por primera vez por Cox y Ellsworth en 1997, quienes lo definieron como el proceso de graficar grandes cantidades de información de manera fidedigna, dadas las limitaciones técnicas de las computadoras de entonces (Mashey, 1998). Lupton (2015) aporta una perspectiva adicional, destaca que big data se refiere a datos que exhiben una mayor variedad que los convencionales, se presentan en volúmenes en constante crecimiento, se generan a una velocidad considerable y adquieren valor al poseer inteligencia intrínseca.

En contrapartida, este trabajo se ha apoyado en la mirada crítica de Boyd y Crawford (2012) para definir big data como un fenómeno cultural, tecnológico y científico —que maximiza la potencia de cálculo y la precisión algorítmica para recopilar, analizar, vincular y comparar grandes conjuntos de datos, y analizarlos a partir de la identificación de otros múltiples patrones. Como consecuencia, las tecnologías de macrodatos y procesamiento computacional de la información pasaron, con su masiva utilización y conocimiento, a vincularse a la

creencia de que los grandes conjuntos de datos ofrecen una forma superior de inteligencia y conocimiento que puede generar ideas que antes eran imposibles.

Es precisamente debido a la promesa de una nueva verdad, objetividad y precisión, vinculada a las técnicas computacionales de procesamiento de datos y al empleo de tecnologías de big data, que a partir de la década de 2010, estas han adquirido un lugar destacado en el ámbito científico, incorporándose de manera definitiva en las metodologías de diversas disciplinas (Navarro y Salazar, 2007). El término ya no se limita al ámbito de los mercados, bancos y estudios de marketing, sino que se integra en pleno a la esfera científica, constituyendo el trasfondo principal de esta investigación.

Con el propósito de elucidar el proceso de introducción del término big data en las ciencias sociales latinoamericanas y evaluar el impacto de la adopción de tecnologías computacionales de procesamiento de información, realicé una investigación exhaustiva en las dos bases de datos más populares en América Latina: Scopus y Scielo. El objetivo fue conocer el campo científico que ha surgido en torno a los datos masivos, así como exponer las especificidades metodológicas relacionadas con este fenómeno y las posibles implicaciones sociales resultantes de este proceso.

En el intento de establecer un correlato entre el descubrimiento, los procesos y las normas científicas de investigación relacionadas con big data y su correspondencia con aspectos presentes y emergentes en una nueva sociedad (computacional), elegí realizar una investigación etnográfica. Esta se basa en técnicas tal como entrevistas semi-estructuradas, observación participante, y grupos focales y seguimientos, los que constituyen los cimientos del proceso de obtención de información. Este enfoque buscó la generación de un conocimiento científico que se ajustara a mi condición de investigadora, incorporando las conexiones transc científicas de la investigación (Knorr-Cetina, 2005), vivenciadas a través de mi experiencia como miembro activo y generador de conocimiento en los estudios constructivistas de ciencia, tecnología y sociedad.

En mi comprensión, el laboratorio científico, la difusión de los resultados investigativos mediante la redacción de artículos y textos académicos, así como la observación del proceso de construcción de sus propios saberes, se perciben como un estudio de caso relevante para comprender las transformaciones en la sociedad. Esto se debe a que entiendo estos componentes como sistemas sociotécnicos intrínsecamente entrelazados (Latour (1992; 2008;

2013). La búsqueda por comprender las particularidades de un laboratorio científico revela las costumbres, principios y arreglos sociales compartidos dentro de la misma sociedad. En última instancia, ciencia, tecnología y sociedad son partes integrantes de una misma construcción social, reflejándose mutuamente, cada uno ofreciendo perspectivas valiosas sobre el otro.

En la fase inicial de esta investigación, mediante un análisis bibliométrico, se evidenció que las publicaciones en ciencias sociales relacionadas con big data en Latinoamérica son relativamente escasas en comparación con el panorama global. Estas manifestaciones se caracterizan por ser textos predominantemente teóricos, priorizando la discusión sobre las posibilidades y limitaciones de la aplicación de big data en sus respectivos campos de investigación o disciplinas afines.

Al examinar las palabras clave, títulos y resúmenes de los textos académicos, se observa que la introducción de la temática big data está asociada a vocabulario del ámbito científico, tales como "investigaciones", "research", "study", "analyse", "information" y "challenge". Este hecho sugiere una mayor concentración de textos de naturaleza teórica en lugar de metodológica. Existe un énfasis en la explicación de conceptos, posibles implicaciones y cambios en la práctica científica a raíz de la incorporación de estas tecnologías.

Asimismo, se vincula con términos y contenidos que giran en torno a la discusión sobre los modelos de sociedad contemporáneos, tales como "data", "digital data", "capitalism", "society information", "new cities". Esto evidencia un debate macropolítico y social asociado a las cuestiones de big data que se relacionan con la digitalidad, nuevas tecnologías, redes sociales, así como estudios relacionados con la modernidad, la juventud, los modos de consumo, la digitalización de la vida y la sociedad de control. Este patrón también se confirmó durante las entrevistas. Al dialogar con los autores seleccionados, constaté mis impresiones de la lectura y análisis de contenido de las publicaciones incluidas en la muestra. Estos autores señalan que la necesidad de abordar y discutir sobre big data surge de la intersección de sus temas con las dimensiones asociadas a la digitalización de la vida, los entornos tecnológicos y la contemporaneidad.

Por lo tanto, se observa una falta de publicaciones que describan resultados de investigaciones o la aplicación de técnicas específicas, en el ámbito de las ciencias sociales. Además, la temática de big data tiende a tratarse en un segundo plano en los textos

analizados, surgiendo como una discusión que orbita alrededor del tema principal de los artículos. Aquellos científicos sociales que se involucran con la temática big data lo hacen en el contexto de trabajar en diversas áreas relacionadas con datos digitales, ciencias de datos o inteligencia artificial. En su mayoría, desempeñan roles como consultores, analistas o facilitadores de metodologías cualitativas incorporadas a proyectos sociales que utilizan procedimientos de ciencia de datos.

A través de las entrevistas y el estudio etnográfico, pude constatar que los pocos científicos sociales latinoamericanos que participan en la investigación basada en macrodatos, lo hacen como integrantes de un proyecto externo (es decir, no como el responsable o coordinador principal) que incorpora big data en alguna fase de su proceso investigativo. Estos profesionales no interactúan directamente con las aproximaciones en tecnologías big data en sus labores de investigación, sino que participan en proyectos que utilizan estos enfoques como parte de su metodología.

También destaco un hallazgo significativo en relación con la evolución o madurez en el uso de tecnologías big data: la disminución del empleo de este término y la introducción del concepto y término "ciencia de datos" en los contextos de dichas investigaciones. Sostuve que la adopción de una nueva expresión para referirse al mismo fenómeno estudiado representa una consolidación en el uso de técnicas y metodologías de procesamiento computacional de información dentro de las ciencias sociales, así como una profundización en la comprensión de los significados reales de estas prácticas.

Considero que la apropiación de este concepto viene acompañada del desarrollo del principio de colaboración y cooperación entre profesionales de diversas áreas del conocimiento, especialmente con ingenieros computacionales, desarrolladores y especialistas técnicos involucrados en estas investigaciones. La coexistencia colaborativa de estos saberes ha posibilitado la inclusión del concepto de ciencia de datos en el ámbito de las investigaciones sociales, reservando el término big data exclusivamente para referirse a ciertos aspectos característicos de los datos o bases de datos que se manejan computacionalmente.

La interacción entre los diversos actores, tanto humanos como no humanos en el ámbito de la investigación social, ocurre a través del contexto investigativo, el flujo de información generado y, sobre todo, alrededor de la problemática estudiada. Este entorno constituye un sistema sociotécnico dinámico, donde los distintos actores se agrupan en comunidades de

especialistas o pequeños grupos afines que, a su vez, se adaptan a las condiciones particulares de su contexto y entorno. Este proceso de incorporación de particularidades mediante la correlación mutua confiere al sistema de investigación en ciencia de datos una dinámica y características únicas.

El entorno tecnológico asociado a la ciencia de datos da forma y establece los límites de este sistema sociotécnico. Elementos clave en esta configuración incluyen la infraestructura de datos, la capacidad computacional para el procesamiento, la capacidad de gobernanza de los datos y el capital humano capaz de manipular datos digitales en variados formatos. Asimismo, el lenguaje y el conocimiento necesarios para operar en sintonía con los tiempos y desafíos involucrados juegan un papel esencial y determinante en este sistema. En este contexto, diversos especialistas actúan como pequeñas comunidades representantes de saberes muy diversos, coexistiendo, estableciendo relaciones y ejerciendo mutua influencia.

A través del seguimiento y análisis etnográfico de un laboratorio de innovación pública con énfasis en ciencia de datos, pude observar transformaciones significativas en la introducción de principios investigativos en el ámbito de los estudios sociales mediante tecnologías big data y ciencia de datos. Estas transformaciones han tenido un impacto considerable en la forma en que los científicos sociales contemporáneos llevan a cabo sus investigaciones. La orientación hacia la investigación aplicada, el diseño investigativo consolidado a partir de los datos disponibles, la promoción de la colaboración e interdisciplinariedad, la confianza en la imprevisibilidad, la disminución de la importancia de la muestra y la imperativa cuantificación y visualización de datos son consecuencias directas de este enfoque investigativo. Estas implicaciones son responsables de remodelar las prácticas y conocimientos generados a partir de esta nueva forma de abordar la investigación social.

La emergencia y consolidación de esta otra práctica científica dentro de nuevos sistemas de investigación se comprenden como elementos cruciales para entender cómo se generan o validan los conocimientos en la sociedad. Se reconoce el hecho de que la ciencia de datos está configurando, representando, un tipo de sociedad (la sociedad computacional) que es valiosa y merece ser observada desde una perspectiva científica. Esta sociedad se caracteriza por ciertos principios tales como el tecno-encanto, el principio de automatización, eficiencia y aceleración de los procesos. Argumento que la digitalización de la vida social y la conectividad a través de internet contribuyen a la instauración de una sociedad que se

desarrolla según los principios de la velocidad, la identidad global, la alienación del trabajo humano detrás de la máquina y una identidad anfibia.

Los cambios observados en la sociedad computacional, especialmente en un mundo post pandémico, están transformando la estructura social en diferentes esferas de la vida, como la economía, la cultura, la política y el ocio. La forma computacional de constituir la sociedad impacta de manera significativa en dinámicas clave de la vida social (Ponce-Tituaña y Lucio-Paredes, 2021). Como era de esperar, el mundo laboral también se ve afectado, incluyendo el ámbito de los científicos sociales, quienes son parte integral de la sociedad.

En el último capítulo, me centré en las características y dimensiones asociadas a este nuevo profesional en ciencias sociales, destacando las características actuales que me permitieron especular sobre las prácticas, actividades y diversas ocupaciones de este tipo de profesional en humanidades. Mientras se especializa, se cualifica y se adapta a las transformaciones de la sociedad computacional, surge y se estructura un nuevo campo disciplinario de estudio y trabajo: las humanidades digitales. Esta disciplina ha cobrado fuerza especialmente a partir de la década de 2010, ofreciendo a los científicos sociales un nuevo campo de trabajo, así como otra posibilidad para analizar y comprender el mundo social contemporáneo que nos rodea.

En síntesis, este estudio ha explorado las transformaciones en la investigación social impulsadas por la introducción de tecnologías big data y ciencia de datos, especialmente en el contexto de un laboratorio de innovación pública. La observación de cambios significativos en las prácticas científicas y la emergencia de nuevas dinámicas en la sociedad computacional, se puede ver en los estudios de la sociedad, ciencia y tecnología. Estos, plantean una reconfiguración de una idea de lo social, que a su vez, representa no sólo la manera como científicos sociales han pasado a representar, entender y estudiar la sociedad, sino que también los ha llevado a adaptarse a una nueva manera de proceder su quehacer científico.

Este trabajo identifica las características del nuevo profesional en ciencias sociales, a la vez que proyecta la aparición del emergente campo disciplinario de las humanidades digitales, las que han surgido como una respuesta a las demandas y desafíos de la sociedad computacional. Espero que esta investigación sirva como un buen ejemplo de caso para comprender el vínculo entre los saberes desarrollados en los laboratorios y las prácticas sociales contemporáneas.

La relevancia de esta investigación para los estudios contemporáneos de la ciencia, tecnología y sociedad (CTS) radica en su capacidad para arrojar luz sobre la intersección entre la innovación tecnológica y las dinámicas sociales en América Latina. Al explorar cómo la ciencia de datos impacta la investigación social y contribuye a la formación de nuevas disciplinas, como las humanidades digitales, este estudio representa un avance significativo en el campo de los estudios CTS en la región.

Reconozco, a su vez, un vasto horizonte de temas por explorar más allá de lo abordado en esta investigación. Considero que la exploración de la digitalización en el ámbito gubernamental (los llamados “datos de gobierno”), así como el análisis de macrodatos en el campo de la administración pública, especialmente en América Latina, constituyen campos fructíferos. Aunque estos temas son un notable complemento a lo estudiado acá, mi impresión es que profundizar en ellos, a pesar de su riqueza, sobrepasaría considerablemente mis conocimientos y el tiempo oficial necesario para esta investigación. Por ello, podrían ser abordados en futuras investigaciones, ya sea por otros científicos o por mí misma en etapas posteriores de mi carrera profesional.

Un interesante tema a profundizar sería un estudio bibliométrico que considere, bajo los mismos parámetros de esta investigación, las publicaciones académicas más recientes, sobretodo a partir de 2021 y 2022, a partir de los términos “ciencia de datos” y “data science” en las ciencias sociales. Esto ayudaría a entender mejor el modo en que conceptos y significados cambian, a medida que la comprensión de “big data” pasa a incluir, o definitivamente reemplazar, a “ciencia de datos”.

Considerando los comprensibles límites de esta investigación, espero que haya contribuido al crecimiento y avance de la disciplina y los estudios CTS latinoamericanos. Al proporcionar un análisis detallado de las transformaciones en la investigación social en el contexto de la sociedad computacional, he aspirado a fomentar una comprensión más profunda de las interacciones entre ciencia, tecnología y sociedad en la era digital, promoviendo así un diálogo crítico y enriquecedor en la comunidad académica.

### Referencias

Adjerid, I., y Kelley, K. (2018). Big Data in Psychology: A framework for Research Advancement. *American Psychologist*, 73(7), 899-917. <https://doi.org/10.1037/amp0000190>

Anderson, C. (23 de junio de 2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired*. <http://www.wired.com/2008/06/pb-theory/>

Antunes, Deborah y Maia, Ari. (2018). Big Data, exploração ubíqua e propaganda dirigida: novas facetas da indústria cultural. *Psicologia USP*. 29. 189-199. <https://doi.org/10.1590/0103-656420170156>

Arcila-Calderón, C., Ortega-Mohedano, F., Jiménez-Amores, J. y Trullenque, S. (2017). Análisis supervisado de sentimientos políticos en español: clasificación en tiempo real de tweets basada en aprendizaje automático. *Profesional de la Información*, 26(5), 973-982. <https://doi.org/10.3145/epi.2017.sep.18>

Arévalo, J. (22 de febrero de 2018). ¿Qué es la minería de textos, cómo funciona y por qué es útil?. *Universo Abierto*.  
<https://universoabierto.org/2018/02/22/que-es-la-mineria-de-textos-como-funciona-y-por-que-es-util/>

Arthur, C. (2017, 21 febrero). What's a zettabyte? By 2015, the internet will know, says Cisco. *the Guardian*.  
<https://www.theguardian.com/technology/blog/2011/jun/29/zettabyte-data-internet-cisco>

Ballivian, A. (29 de abril de 2016). El uso de Big Data para estadísticas oficiales y los objetivos de desarrollo sostenible. [Ponencia]. Seminario *La Revolución del Big Data en Estudios Sociales*, El Colegio de México, México.

Bastin, G. y Tubaro, P. (2018). Le moment Big Data des sciences sociales. *Revue Française de Sociologie*, 59(3), 375-394. <https://doi.org/10.3917/rfs.593.0375>

Bearman, P. (2015). Big Data and Historical Social Science. *Big Data & Society*, 2(2).  
<https://doi.org/10.1177/2053951715612497>

Bordalejo, B., O'Donnell, D. P., Gonzalez, E., Risam, R. y Spence, P. (2014). Global Outlook: Digital Humanities: Promoting Digital Humanities Research Across disciplines, Regions, and Cultures. Conferencia. [Conferencia]. *Digital Humanities 2014*, École Polytechnique Fédéral, Suiza.

Boyd, D. y Crawford, K. (2011). Six Provocations for Big Data [ponencia]. Simposio *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*. Instituto de Internet de Oxford, Inglaterra.

Bunge, M. (1997). *Social science under debate: a philosophical perspective*. University of Toronto Press.

Burkholder, L. (1992). *Philosophy and the Computer, Boulder, San Francisco, and Oxford*. Westview.

Brady, H. (2019). The Challenge of Big Data and Data Science. *Annual Review of political science*, 22(1), 297-323. <https://doi.org/10.1146/annurev-polisci-090216-023229>

Calderón, F. (2004). *¿Es sostenible la globalización en América Latina?* (Vol.1). Fondo de Cultura Económica.

Calderón, F. y Szmukler, A. (2004). Cultura de igualdad, deliberación y desarrollo humano en *La construcción social de los derechos y la cuestión del desarrollo: antología esencial*. CLACSO. <https://doi.org/10.2307/j.ctv253f4hk>

Callon, M., Law, J. y Rip, A. (1986). *Mapeo de la dinámica de la ciencia y la tecnología: Sociología de la ciencia en el mundo real*. Basingstoke: Macmillan.

Castells, M. (1996). *La era de la información. Economía, sociedad y cultura*. (Vol. 1). Siglo XXI

Castillo-Esparcia, A., Álvarez-Nobell, A., y Barroso, M.B. (2021). ‘Issues’ y ‘big data’ en la gestión de relaciones públicas. El caso de la implementación del nuevo sistema de tratamiento de residuos “Recuperando Valor” en Córdoba, Argentina. *Tripodos. Blanquerna School of Communication and International Relations-URL* (45), 73-87. <https://raco.cat/index.php/Tripodos/article/view/363334>.

Chavan, T. y Akarte, S.P. (2014). Opportunities and Challenges of Big Data in Economics Research and Enterprises. *International Journal of Computer Science and Mobile Computing*, 3(4), 1155-1161.

Chavez, H., Albornoz, M. B. y Martín, F. (2022). ‘Big data’ Research: A Bibliometric Analysis of the Scopus Database, 2009–2019. *Journal of Scientometric Research*, 11(1), 64–78. <https://doi.org/10.5530/jscires.11.1.7>

Collins C., Kanyifechukwu J. y Bisallah, H. (2021). Big Data And Analytics Implementation In Tertiary Institutions To Predict Students Performance In Nigeria. *ScienceOpen Preprints*. <https://doi.org/10.48550/arXiv.2207.14677>

Comisión Económica para América Latina y el Caribe (CEPAL) (2021). *Panorama social de América Latina 2020*. Naciones Unidas. [https://www.cepal.org/sites/default/files/publication/files/46687/S2100150\\_es.pdf](https://www.cepal.org/sites/default/files/publication/files/46687/S2100150_es.pdf)

Comisión Gulbenkian para la reestructuración de las ciencias sociales (1996). *Abrir las ciencias sociales*. Siglo xxi Editores.

Conte, R., Gilbert, N., Bonelli, G., Cioffi-Revilla, C., Deffuant, G., Kertész, J., Loreto, V., Moat, H. S., Nadal, J., Sánchez, Á., Nowak, A., Flache, A., Miguel, M. S., y Helbing, D. (2012). Manifesto of computational social science. *European Physical Journal-special Topics*, 214(1), 325–346. <https://doi.org/10.1140/epjst/e2012-01697-8>

Cuadros Garrido, M.E (2019). La protección de los derechos fundamentales de la persona trabajadora ante la utilización de GPS: ¿reformulación o continuidad? *Lan Harremanak - Revista De Relaciones Laborales*, (42). <https://doi.org/10.1387/lan-harremanak.21087>

Devins, C., Felin, T., Kauffman, S. y Koppl, R. (2017). The Law and Big Data. *Cornell Journal of Law and Public Policy*, 27(2), 357-413.

DiMaggio, P. (2015). Adapting computational text analysis to social science (and vice versa). *Big Data & Society*, 2(2). <https://doi.org/10.1177/2053951715602908>

Domingues, J. M. (2015) Proyecciones de la teoría sociológica en América Latina: descripción, análisis y diagnóstico de la modernidad. *POLIS, Revista Latinoamericana*, 14(41), 97-100. <http://dx.doi.org/10.4067/S0718-65682015000200007>.

Estrada, A.H., Martínez, B.A., Pérez, E.G., Figueroa, J.L., Martín, M.N., Morales, G.M., Rodríguez, M., Sánchez, M.D., Armendáriz, M.M., Sáez, M.Á., Gómez, I.P., Santos, J.M., García, Á.D., Pinto, A.R., y Fuente, J.I. (2020). *Big Data en Educación III: recomendaciones personalizadas de actuación ante estudiantes en riesgo de abandono* [Proyecto número 25, Convocatoria 2018/2019] Universidad Complutense de Madrid. <https://hdl.handle.net/20.500.14352/9442>

Ferreira, M. (2001). Más allá del Laboratorio. La Antropología del Conocimiento. Científico como apuesta Metodológica, *Política y Sociedad*, 37, 105-126.

Flores Vivar, J. M. (2018). Algoritmos, aplicaciones y big data, nuevos paradigmas en el proceso de comunicación y de enseñanza-aprendizaje del periodismo de datos. *Revista de comunicación*, 17(2), 268-291. <http://dx.doi.org/10.26441/RC17.2-2018-A12>

Follari, R. (2015). Las ciencias sociales en la encrucijada actual. *POLIS, Revista Latinoamericana*, 14(41), 147-157. <http://polis.revues.org/11162>

Garretón, M. (2000) *La sociedad en la que vivi(re)mos. Introducción sociológica al cambio de siglo*. LOM.

Gauthier, E. (1998). Bibliometric Analysis of Scientific and Technological Research: A User's Guide to the Methodology. *Science and Technology Redesign Project Statistics Canada*, 8.

Gómez Barrera, J.C. y Arango Tobón, M.A. (2022). Psicometría, subjetividad y redes sociales: el estudio del lenguaje en Facebook. *Revista de Psicología Universidad de Antioquia*, 13(2), 1-27. <https://doi.org/10.17533/udea.rp.e345913>

Gómez-Barroso, J.-L. y Feijóo-González, C. (2013). Información personal: la nueva moneda de la economía digital. *Profesional De La información Information Professional*, 22(4), 290–297. <https://doi.org/10.3145/epi.2013.jul.03>

Grigar, Dene. (2013). *Debates in the Digital Humanities* by Matthew K. Gold. Leonardo. 46. 186-187. [10.2307/23468163](https://doi.org/10.2307/23468163).

Gutiérrez, C. (2018). Datos, ciencia y sociedad en el siglo XXI. *Revista BITS de Ciencia*, (16), 10-17.

Gurstein, M (2011). Open data: Empowering the Empowered or Effective Data Use for Everyone?. *First Monday*, 16 (2). <https://doi.org/10.5210/fm.v16i2.3316>

Han, J., Kamber, M. y Pei, J.(2011). *Data Mining: Concepts and Techniques*. (3ra ed). Morgan Kaufmann.

Hernández, J.C. (2020). Decisiones algorítmicas de perfilado: Régimen y garantías jurídicas. *Revista Española de Derecho Administrativo*, (203), 281-322.

Jara Fuentealba, N. (2021). El derecho de la propiedad sobre los datos. *Revista Chilena de derecho privado*, (temático), 101-142. <https://dx.doi.org/10.4067/S0718-80722021000300101>

Kitchin, R. (2013). Big Data and Human Geography: Opportunities, Challenges and Risks. *Dialogues in human geography*, 3(3), 262-267. <https://doi.org/10.1177/2043820613513388>

Knorr-Cetina, K., Strasser, H. y Zillian, H. (1975). *Determinants and Controls of Scientific Development*. D. Reidel Publishing Company.

Knorr-Cetina, Karin (1981). *Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature Of Science*. (1ra ed.) Pergamon.

Knorr-Cetina, K. (1995) Laboratory Studies: The Cultural Approach to the Study of Science. En Jasanoff, S. et al. (eds.), *Handbook of Science and Technology Studies*. Sage.

Knorr-Cetina, Karin (2005). *La fabricación del conocimiento: un ensayo sobre el carácter constructivista y contextual de la ciencia*. Editorial Bernal.  
<http://ridaa.unq.edu.ar/handle/20.500.11807/595>

Kurniasih, Nia. (2021). *DIGITAL HUMANITIES: A Primer for Scholars and Students*. Jurnal Sositologi. 20. 132-136. 10.5614/sostek.itbj.2021.20.1.13.

Laboratorio de innovación pública de la Escuela de Gobierno [Gob\_Lab] (s.f). *Ética de datos*. <https://goblab.uai.cl/etica-de-datos/>

Lansley, G., de Smith, M., Goodchild, M. y Longley, P. (2018). Big Data and Geospatial Analysis en *Geospatial Analysis: A comprehensive guide to principles, techniques and software tools*. (6ta ed.). Winchelsea. <https://doi.org/10.48550/arXiv.1902.06672>

Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford University.

Latour, B. (2011) *Ciência em ação: como seguir cientistas e engenheiros sociedade afora*. (2da ed). Unesp.

Latour, B. (2008) *Reensamblar lo social: Una introducción a la teoría del actor-red*. Manantial.

Latour, B. y Woolgar, S. (1995) *La vida en el laboratorio. La construcción de los hechos científicos*. Alianza.

Liu, Z., Guo, H. y Wang C. (7-8 de julio de 2016). IOP Conference Series: Earth and Environmental Science 46. Beijing, China.  
<https://iopscience.iop.org/article/10.1088/1755-1315/46/1/012058>

- London School of Economics [LSE] Benoit, K. y Cukier, K. (19 de febrero de 2015). *The challenges of Big Data for the Social Sciences*. [Archivo de Vídeo]. Youtube.  
[https://www.youtube.com/watch?v=IG9RfpGTmJk&ab\\_channel=LSE](https://www.youtube.com/watch?v=IG9RfpGTmJk&ab_channel=LSE)
- López Espinosa, J. (2022). Ventajas de la utilización de big data en el proceso auditor, 2018. *Revista chilena de la administración del estado*, (1).  
<https://doi.org/10.57211/revista.v1i01.34>
- Lucía Megías, J.M. (2003). La “Informática Humanística: Notas Volanderas Desde El ámbito hispánico». *INCIPIT*, 23, 91-114.  
<https://ojs.iibicrit-conicet.gob.ar/index.php/incipit/article/view/436>.
- Madyatmadja, E., Sembiring, D., Angin, S., Ferdy, D. y Andry, J. (2021). Big Data in Educational Institutions using RapidMiner to Predict Learning Effectiveness. *Journal of Computer Science*, 17(4), 403-413. <https://doi.org/10.3844/jcssp.2021.403.413>
- Maestro-Cano, I. (2019). Sobre las limitaciones de Big Data en ciencias sociales. *Sociología y tecnociencia*, 9(2). <https://doi.org/10.24197/st.2.2019.77-98>
- Martínez-Uribe, L. (2022). La sociología a través de sus publicaciones en revistas de impacto mediante el uso de big data. *EMPIRIA. Revista de metodología de ciencias sociales* (53), 53-88. <https://doi.org/10.5944/empiria.53.2022.32612>
- Mauss, Marcel. Ensaio sobre a dádiva. in: Mauss, M. *Sociologia e Antropologia*. SP, Cosac Naif, 2003
- Mignolo, W. (2008) La opción decolonial: desprendimiento y apertura: Un manifiesto y un caso. *Tabula Rasa*, 8, 243 – 281. <https://doi.org/10.25058/20112742.331>
- Milan, S., y Gutiérrez, M. (2015). Citizens’ Media Meets Big Data: The Emergence of Data Activism. *MEDIACIONES*, 11(14), 120–133.  
<https://doi.org/10.26620/uniminuto.mediaciones.11.14.2015.120-133>
- Miller, D. y Slater, D. (2000) *The internet: An Ethnographic Approach*. (1era ed). Berg.

Miller, D. (2012) *The Digital and the Human: A Prospectus for Digital Anthropology*. En H. Horst y D. Miller (Eds). *Digital Antropology* (pp. 3-36). Berg.

<https://doi.org/10.4324/9781003085201-2>

Mendoza, M. (2001). En la tribu de los científicos. *POLIS, Revista Latinoamericana*, 1(1), 0.

Monleon-Getino A. (2015). El impacto del Big-data en la Sociedad de la Información. Significado y utilidad". *Historia y Comunicación Social*, 20(2), 427-445.

[https://doi.org/10.5209/rev\\_HICS.2015.v20.n2.51392](https://doi.org/10.5209/rev_HICS.2015.v20.n2.51392)

Monroe, B. L. (2013). The Five Vs of Big Data Political Science Introduction to the Virtual Issue on Big Data in Political Science Political Analysis. *Political Analysis*, 21(V5), 1-9. <https://doi.org/10.1017/S1047198700014315>

Mulkay, M. (1979). *Science and the Sociology of Knowledge*. (1era ed.). Routledge. <https://doi.org/10.4324/9781315763408>

Mützel, S. (2015). Facing Big Data: Making Sociology Relevant. *Big Data & Society*, 2(2). <https://doi.org/10.1177/2053951715599179>

Nascimento, L. F. (2016). A Sociologia Digital: um desafio para o século XXI. *Sociologias*, 18(41), 216–241. <https://doi.org/10.1590/15174522-018004111>

Navarro, L. & Salazar, J. (2007) Análisis de redes sociales aplicado a redes de investigación en ciencia y tecnología. *Síntesis Tecnológica*, 3(2), 69-86.

<https://doi.org/10.4206/sint.tecnol.2007.v3n2-03>

Organización para la Cooperación y el Desarrollo Económico. (1996). *The knowledge-based Economy*. OECD, Paris.

Pérez-Matos N. (2002). La bibliografía, bibliometría y las ciencias afines. *ACIMED*, 10(3), 1-2.

Puebla, J. G. (2018). Big Data y nuevas geografías: la huella digital de las actividades humanas. *Documents d'anàlisi geogràfica*, 64(2), 195-217.

<https://doi.org/10.5565/rev/dag.526>

- Quijano, A. (2014). Colonialidad del poder, eurocentrismo y América Latina. En: *Cuestiones y horizontes: de la dependencia histórico-estructural a la colonialidad/descolonialidad del poder*. CLACSO.
- Raban, D y Gordon A. (2020). The Evolution of Data Science and Big Data Research: A Bibliometric Analysis. *Scientometrics*, 122(3), 1563-1581.  
<https://doi.org/10.1007/s11192-020-03371-2>
- Ribes, D. (2018). STS, Meet Data Science, Once Again. *Science, Technology, & Human Values*, 44(3). <https://doi.org/10.1177/016224391879>
- Ruiz, M.A., Gómez-Zúñiga, B., Encuentra, E.H. y Pousada, M. (2015). Big data y Psicología: ¿una oportunidad para el Internet de las personas? *Aloma*, 33(2)  
<https://doi.org/10.51698/aloma.2015.33.2.21-29>
- Sabino, A.-B., Reis-Martins, P., y Carranza-Infante, M. (2020). Experiencias y retos del uso de datos de aplicaciones móviles para la movilidad urbana. *Revista de Arquitectura (Bogotá)*, 22(1), 82–93. <https://doi.org/10.14718/RevArq.2020.3039>
- Saisó, E. (2008) Revista Digital Universitaria: Facing the Challenges of Digital Publishing in Mexico. *Storicamente*, 4(37). 10.1473/stor301
- Salazar Argonza, J. (2016). Big Data en la educación, *Revista Digital Universitaria*, 17(1), 1-16. <http://www.revista.unam.mx/vol.17/num1/art06/index.html>
- Schwab, K. (2016). *La cuarta Revolución Industrial en Foro Económico Mundial*. Penguin Random House.
- Slota, S. C., Hoffman, A. S., Ribes, D., & Bowker, G. C. (2020). Prospecting (in) the data sciences. *Big Data & Society*, 7(1). <https://doi.org/10.1177/2053951720906849>
- Spence, P. y González-Blanco García, E. (2014). A Historical Perspective on the Digital Humanities in Spain. *H-Soz-Kult*, 1-19. <https://doi.org/10.3145/epi.2020.nov.01>
- Serrano Pérez, M.M. (2015). Big Data o la acumulación masiva de datos sanitarios: derechos en riesgo en el marco de la sociedad digital. *Revista Derecho y Salud*, 25(1), 51-64.  
<https://dialnet.unirioja.es/servlet/articulo?codigo=5635389>

Shorey, S. y Howard, P. (2016). Automation, big data and politics: a research review. *International Journal of Communication*, 10, 5032–5055.

Téllez Carvajal, E. (2020). Análisis documental sobre el tema del Big Data y su impacto en los derechos humanos. *Derecho PUCP*, (84), 155-188.  
<https://dx.doi.org/10.18800/derechopucp.202001.006>

Tinati, R., Halford, S., Carr, L. y Pope, C. (2014). Big data: Methodological challenges and approaches for sociological analysis. *Sociology*, 48(4), 663-681.  
<https://doi.org/10.1177/0038038513511561>

Varian, H.R. (2014). Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives*, 28(2), 3-28. DOI: 10.1257/jep.28.2.3

Vivar, J.M. (2018). Algoritmos, aplicaciones y Big data, nuevos paradigmas en el proceso de comunicación y de enseñanza-aprendizaje del periodismo de datos. *Revista de Comunicación*, 17(2), 268-291. <https://doi.org/10.26441/RC17.2-2018-A12>

Wang B. y Li, X. (2017). Big Data, Platform Economy and Market Competition: A Preliminary Construction of Plan-Oriented Market Economy System in the Information Era. *World Review of Political Economy*, 8(2):138-161.  
<https://doi.org/10.13169/worlrevipoliecon.8.2.0138>

Wang, LuQuan. (2020). Research on the Application of Big Data Statistical Analysis in the Field of Economic Management. *Journal of Physics: Conference Series*, 1533. 042010.  
<https://doi.org/10.1088/1742-6596/1533/4/042010>

Wang, J., Wu, Y., Yen, N., Guo, S. y Cheng, Z. (2016). Big Data Analytics for Emergency Communication Networks: A Survey. *IEEE Communications Surveys & Tutorials*, 18(3), 1758-1778. <https://doi.org/10.1109/COMST.2016.2540004>

Wiencierz, C. y Röttger, U. (2017). The Use of Big Data in Corporate Communication. *Corporate Communications: An International Journal*, 22(3), 258-272.  
<https://doi.org/10.1108/CCIJ-02-2016-0015>

Walsh, John & Cobb, Peter & Fremery, Wayne & Golub, Koraljka & Keah, Humphrey & Kim, Jeonghyun & Kiplang'at, Joseph & Liu, Ying-Hsang & Mahony, Simon & Oh, Sam & Sula, Chris Alen & Underwood, Ted & Wang, Xiaoguang. (2021). Digital

humanities in the iSchool. Journal of the Association for Information Science and Technology. 73. 10.1002/asi.24535.

Wright Mills, C. (2003). *La imaginación sociológica*. Fondo de Cultura Económica.

## **Anexos**

### **I. Glosario**

Este glosario fue construido a 4 manos: a partir de la síntesis de los conceptos trabajados con los entrevistados, mi propio entendimiento y capacidad de resumirlos. Fue revisado y complementado a través de la cooperación de un Ingeniero Matemático y programador, mi marido. Contamos con el auxilio de nuestro querido asistente de investigación conceptual, ChatGPT 3.5.

#### **Algoritmo**

Un algoritmo es un conjunto de pasos o reglas definidas y ordenadas que resuelven un problema o realizan una tarea específica. Estos pasos son diseñados para ser ejecutados de manera precisa y eficiente por una computadora o por un agente humano. Los algoritmos son fundamentales en ciencias de la computación y matemáticas, proporcionando un método sistemático para abordar y resolver problemas.

#### **Cloud-Computing**

La computación en la nube se refiere a la entrega de servicios de computación a través de Internet, permitiendo a las organizaciones utilizar recursos como servidores, almacenamiento, bases de datos y software a través de proveedores

de servicios en la nube en lugar de mantenerlos localmente. Aspectos clave incluyen la escalabilidad, que permite ajustar los recursos según la demanda; el modelo de pago por uso, evitando inversiones costosas en hardware; accesibilidad desde cualquier lugar con conexión a Internet; respaldo y recuperación de datos automatizados; actualizaciones y mantenimiento gestionados por proveedores; y medidas de seguridad avanzadas, siendo la seguridad una responsabilidad compartida entre proveedor y cliente. Este enfoque ofrece flexibilidad, eficiencia y ahorro de costos a las organizaciones.

#### **Data Mining**

La minería de datos o Data Mining es un proceso de exploración y análisis de extensos conjuntos de datos con el

propósito de descubrir patrones, tendencias, relaciones y conocimientos ocultos. Utilizando técnicas de estadística, aprendizaje automático, inteligencia artificial y análisis de datos, la minería de datos extrae información valiosa de datos brutos. Las etapas clave incluyen la recopilación de datos de diversas fuentes, el preprocesamiento para limpiar y preparar los datos, la exploración para comprender su naturaleza, la selección de algoritmos específicos, la construcción y evaluación de modelos, la interpretación de resultados y, finalmente, el despliegue del modelo para su uso continuo. Este proceso se aplica en campos como el marketing, la medicina, la banca y la ciencia, convirtiendo datos en información accionable para la toma de decisiones informadas.

### **Data Quality**

Calidad de los datos (Data Quality) se refiere al proceso de elevar la calidad de la base de datos de entrada en un proyecto. Se refiere a la evaluación de la precisión, confiabilidad y adecuación de la información utilizada en un proyecto o análisis. Este concepto abarca criterios como exactitud, completitud, consistencia, actualidad, concordancia e integridad referencial. Asegurar la calidad de los datos es esencial en la ciencia de datos para obtener resultados precisos y tomar decisiones fundamentadas. Esto implica procesos como limpieza, corrección de errores y garantía de que los datos sean representativos, contribuyendo a la solidez y confiabilidad de los resultados y decisiones derivadas de ellos.

### **GPU**

Una GPU, o Unidad de Procesamiento Gráfico, es un tipo de procesador diseñado principalmente para gestionar y acelerar el procesamiento de gráficos y datos visuales en una computadora. Optimizadas para tareas gráficas como renderización de imágenes y generación de gráficos en 3D, las GPUs se destacan por su capacidad de procesamiento paralelo, ejecutando múltiples tareas simultáneamente. Con un gran número de núcleos de procesamiento, memoria dedicada (VRAM) para acceso rápido a datos gráficos, y la capacidad de realizar cómputo de propósito general (GPGPU), las GPUs modernas son esenciales para mejorar el rendimiento no sólo en aplicaciones gráficas y multimedia, sino también en cálculos científicos y de alto rendimiento en diversas disciplinas. Su versatilidad las hace fundamentales en el panorama de las computadoras modernas.

### **Entrenamiento supervisado**

El entrenamiento supervisado es una técnica de aprendizaje automático en la que un modelo se entrena utilizando un conjunto de datos que contiene ejemplos de entrada junto con sus correspondientes salidas deseadas. En este proceso, el algoritmo ajusta sus parámetros y características internas a partir de la relación entre las entradas y las salidas conocidas. Durante la fase de entrenamiento, el modelo busca aprender patrones y características relevantes para realizar predicciones o clasificaciones precisas cuando se le presenta nueva información.

La "supervisión" proviene de la existencia de un conjunto de datos de entrenamiento

etiquetado, donde cada ejemplo tiene una etiqueta o salida conocida. El objetivo es que el modelo aprenda a generalizar a partir de estos ejemplos, de modo que, cuando se le presente una nueva entrada no vista durante el entrenamiento, pueda realizar predicciones precisas. El proceso de entrenamiento supervisado se utiliza comúnmente en tareas como clasificación y regresión, donde el modelo se ajusta para prever la clase o el valor de salida correcto en función de las entradas dadas.

### **Entrenamiento no supervisado**

En el entrenamiento no supervisado, un modelo se entrena sin utilizar conjuntos de datos que contengan salidas deseadas. En lugar de seguir etiquetas predefinidas, el algoritmo explora patrones y estructuras inherentes en los datos para descubrir información oculta, relaciones intrínsecas o agrupaciones naturales. Este enfoque es útil para explorar conjuntos de datos sin etiquetas claras y descubrir patrones subyacentes sin la necesidad de prever salidas específicas. Técnicas comunes incluyen la clusterización y la reducción de dimensionalidad.

### **Ingeniero de machine learning:**

Un ingeniero de machine learning es un profesional especializado en la ingeniería de sistemas y aplicaciones que emplean algoritmos de aprendizaje automático (ML). Este campo, perteneciente a la inteligencia artificial, se centra en desarrollar algoritmos y modelos que permiten a las computadoras aprender y mejorar su rendimiento a partir de datos y experiencia. Las responsabilidades clave de un ingeniero de machine learning incluyen la recopilación y preparación de

datos, la selección de algoritmos y modelos, el entrenamiento de modelos, la evaluación y validación del rendimiento, el desarrollo de software que incorpora modelos de ML, la optimización y escalabilidad de sistemas, el mantenimiento continuo, y la colaboración interdisciplinaria con científicos de datos y otros profesionales. Este rol requiere habilidades en programación, matemáticas, estadísticas y conocimientos avanzados en algoritmos de machine learning para diseñar soluciones efectivas y prácticas.

### **Inteligencia Artificial**

La Inteligencia Artificial (IA) es un campo de la informática que se centra en el desarrollo de sistemas y algoritmos capaces de realizar tareas que, normalmente, requieren inteligencia humana. Estas tareas incluyen el aprendizaje, razonamiento, percepción visual, reconocimiento de voz y toma de decisiones. La IA busca imitar la capacidad cognitiva humana para resolver problemas complejos y adaptarse a nuevas situaciones. Su aplicación abarca desde asistentes virtuales y chatbots hasta sistemas avanzados de diagnóstico médico y vehículos autónomos. La IA se basa en algoritmos de aprendizaje automático y redes neuronales, y su evolución continúa transformando diversas industrias.

### **Learning Analytics**

La analítica del aprendizaje es un campo que emplea técnicas de análisis de datos y estadísticas para recopilar, procesar y analizar información relacionada con el proceso educativo. Su objetivo es obtener insights valiosos que mejoren la enseñanza y el aprendizaje. Entre sus aspectos clave se encuentran la recopilación de datos

desde diversas fuentes, el procesamiento utilizando técnicas analíticas, la visualización de resultados comprensibles, la toma de decisiones informadas en la enseñanza, la personalización del aprendizaje, la mejora de la retención y rendimiento estudiantil, y la identificación temprana de problemas. Busca mejorar la educación respetando la privacidad y seguridad de los datos estudiantiles, cumpliendo con regulaciones y políticas de privacidad.

### **Modelo/modelo matemático**

Un modelo matemático es una representación abstracta y estructurada de un sistema o fenómeno de la realidad utilizando conceptos y ecuaciones matemáticas. Estos modelos capturan las relaciones y comportamientos clave del sistema, permitiendo realizar análisis, predicciones o simulaciones. Su aplicación se extiende a diversas disciplinas, como la física, la economía y la ingeniería. Los modelos matemáticos pueden ser expresados mediante ecuaciones diferenciales, algebraicas o probabilísticas, y su utilidad radica en proporcionar una comprensión cuantitativa y formal del sistema en cuestión, facilitando la toma de decisiones y la resolución de problemas complejos.

### **Programadores de back/front-end**

El programador de front-end se ocupa de la parte visible y la experiencia del usuario, mientras que el back-end se encarga de la lógica, el procesamiento de datos y la gestión de la infraestructura. Ambas partes son esenciales para el desarrollo exitoso de aplicaciones web y trabajan juntas para ofrecer una experiencia de usuario integral. La

diferencia principal entre el back-end y el front-end en el desarrollo de software se centra en las áreas de enfoque y las responsabilidades. Ambas partes colaboran para crear aplicaciones web y sistemas de software completos, pero sus roles y tareas son distintos.

### **Script de programación**

Un script de programación es un conjunto de instrucciones o comandos escritos en un lenguaje de programación que realiza tareas específicas. A diferencia de programas completos, los scripts son más ligeros y se ejecutan sin necesidad de compilación, interpretándose directamente por un intérprete o ambiente de ejecución. Estos scripts son frecuentemente utilizados para automatizar procesos, realizar tareas repetitivas o facilitar la interacción con sistemas y aplicaciones, proporcionando una forma eficiente de implementar funcionalidades sin la complejidad de un programa completo.

### **Text mining**

Minería de Texto, en español, es un campo de la ciencia de datos y la inteligencia artificial que se enfoca en extraer información significativa y conocimiento útil a partir de textos no estructurados. Este tipo de textos abarca documentos, correos electrónicos, publicaciones en redes sociales, noticias, artículos, reseñas de productos, entre otros. El proceso de Text Mining implica etapas clave, como la recopilación de datos, el preprocesamiento de texto que incluye la eliminación de elementos no deseados y la normalización, la tokenización para dividir el texto en unidades más pequeñas, el análisis de frecuencia para identificar

palabras clave, la extracción de información específica, el análisis de sentimiento, el modelado de temas, y la visualización de datos para facilitar la interpretación. Se aplica en diversas áreas, como la clasificación de documentos, la recuperación de información, la recomendación de contenido, la detección

de fraudes, la vigilancia de redes sociales, la extracción de información médica, y la toma de decisiones basada en datos textuales. Este enfoque convierte datos no estructurados en información valiosa para comprender patrones, tendencias y conocimientos ocultos en grandes volúmenes de texto.

## II. Tablas de referencia por disciplina

**Tabla Historia Referencia**

Historia	Título Texto	Palabras Claves	Referencia	DOI
1	Facing Big Data: Making sociology relevant	Big Data, sociology, computational social science, topic modeling, textual analysis, data construction	Mützel, S. (2015). Facing Big Data: Making sociology relevant. <i>Big Data &amp; Society</i> , 2(2). <a href="https://doi.org/10.1177/2053951715599179">https://doi.org/10.1177/2053951715599179</a>	<a href="https://journals.sagepub.com/doi/10.1177/2053951715599179">https://journals.sagepub.com/doi/10.1177/2053951715599179</a>
2	Big Data and historical social science	Big Data, historical social science	Bearman, P. (2015). Big Data and historical social science. <i>Big Data &amp; Society</i> , 2(2).	<a href="#">Big Data and historical social science - Peter Bearman, 2015</a>
3	Challenges of Big Data in History	collaborative social change big data history	Manning, P. (2013). Challenges of Big Data in History. In: <i>Big Data in History</i> . Palgrave Pivot, London.	<a href="https://doi.org/10.1057/9781137378972_1">https://doi.org/10.1057/9781137378972_1</a>
4	Big Data for Global History	Big data Global History Digital humanities	Eijnatten, J. V., Pieters, T., & Verheul, J. (2013). Big data for global history: The transformative promise of digital humanities.	<a href="https://dspace.library.uu.nl/handle/1874/288115">https://dspace.library.uu.nl/handle/1874/288115</a>

			BMGN-Low Countries Historical Review, 128(4), 55-77.	
--	--	--	--	--

<b>Economía</b>	<b>Título Texto</b>	<b>Palabras Claves</b>	<b>Referência</b>	<b>DOI</b>
1	Big Data: New Tricks for Econometrics	big data Econometrics	Varian, H.R. (2014). Big Data: New Tricks for Econometrics. Journal of Economic Perspectives, 28, 3-28.	<a href="https://doi.org/10.1257/JEP.28.2.3">DOI:10.1257/JEP.28.2.3</a>
2	Opportunities and Challenges of Big Data in Economics Research and Enterprises	big data Econometrics	Chavan, T., & Akarte, S.P. (2014). Opportunities and Challenges of Big Data in Economics Research and Enterprises.	<a href="https://www.ijcsmc.com/docs/papers/April2014/V3I4201499b76.pdf">https://www.ijcsmc.com/docs/papers/April2014/V3I4201499b76.pdf</a>
3	Research on the Application of Big Data Statistical	Economic management	LuQuan Wang 2020 J. Phys.: Conf. Ser. 1533	<a href="https://iopscience.iop.org/">https://iopscience.iop.org/</a>

	Analysis in the Field of Economic Management	t, Big data, Statistical analysis	042010	<a href="https://doi.org/article/10.1088/1742-6596/1533/4/042010/pdf">org/article/10.1088/1742-6596/1533/4/042010/pdf</a>
4	Review of Big Data Applications in Finance and Economics	Big data (ICTs) data economics	Basdas, U., & Esen, M.F. (2021). Review of Big Data Applications in Finance and Economics	<a href="https://doi.org/10.4018/978-1-7998-3053-5.ch010">DOI:10.4018/978-1-7998-3053-5.ch010</a>
5	Research on the Construction of University Campus Economic Management System Based on the Concept of Big Data	-	Yan Fang, Qiao Wang, "Research on the Construction of University Campus Economic Management System Based on the Concept of Big Data", Mathematical Problems in Engineering, vol. 2022, Article ID 2335221, 9 pages, 2022.	<a href="https://doi.org/10.1155/2022/2335221">https://doi.org/10.1155/2022/2335221</a>

### Tabla Psicología Referencia

Psicología	Título Texto	Palabras Claves	Referência	DOI
1	Big data como objeto de estudio y método para la investigación empírica en sociología y psicología social	Big data Metodologías Sociales Investigación Psicología	Becerra, G. (2018). Big data como objeto de estudio y método para la investigación empírica en	<a href="https://47jaiio.sadio.org.ar/sites/default/files/STS-13.pdf">https://47jaiio.sadio.org.ar/sites/default/files/STS-13.pdf</a>

		social	sociología y psicología social.	
2	Big Data y Psicología: ¿una oportunidad para el Internet de las Personas?	big data; smartphone; app; EMI; TIC	Ruiz, M.A., Gómez-Zúñiga, B., Encuentra, E.H., & Pousada, M. (2015). Big data y Psicología: ¿una oportunidad para el Internet de las personas?	<a href="https://doi.org/10.51698/aloma.2015.33.2.21-29">https://doi.org/10.51698/aloma.2015.33.2.21-29</a>
3	Psicometría, subjetividad y redes sociales: el estudio del lenguaje en Facebook	Psychometry, Facebook, Subjectivity, Societies of Control, Behavior Assessment.	Gómez Barrera, J.C., & Arango Tobón, M.A. (2022). Psicometría, subjetividad y redes sociales: el estudio del lenguaje en Facebook. Revista de Psicología Universidad de Antioquia.	<b>DOI:</b> <a href="https://doi.org/10.17533/udea.rpe.345913">https://doi.org/10.17533/udea.rpe.345913</a>
4	Big data in psychology: A framework for research advancement.	Big data research psychology	Adjerid, I., & Kelley, K. (2018). Big data in psychology: A framework for research advancement. American Psychologist, 73(7), 899.	<a href="https://doi.org/10.1037/amp0000190">https://doi.org/10.1037/amp0000190</a>

Tabla Educación Referencia

Educación	Título Texto	Palabras Claves	Referencia	DOI
1	Big Data en la educación	big data, métodos educativos, TIC, aprendizaje adaptativo, EBC, aula invertida, aprendizaje combinado, gamificación, aprendizaje Móvil, analítica del aprendizaje, Hadoop, SGBD.	SALAZAR ARGONZA, Javier, "Big Data en la educación", Revista Digital Universitaria, 1 de enero de 2016, Vol. 17, Núm. 1. Disponible en Internet: < <a href="http://www.revista.unam.mx/vol.17/num1/art06/index.html">http://www.revista.unam.mx/vol.17/num1/art06/index.html</a> > ISSN: 1607-6079	< <a href="http://www.revista.unam.mx/vol.17/num1/art06/index.html">http://www.revista.unam.mx/vol.17/num1/art06/index.html</a> >
2	Big Data en Educación III: recomendaciones personalizadas de actuación ante estudiantes en riesgo de abandono	Big data Educacion Riesgo abandono escolar	Estrada, A.H., Martínez, B.A., Pérez, E.G., & Fuente, J.I. (2020). Big Data en Educación III: recomendaciones personalizadas de actuación ante estudiantes en riesgo de abandono.	<a href="#">225 Título del Proyecto: Big Data en Educación III</a>
3	Big Data in Educational Institutions using RapidMiner to Predict Learning Effectiveness	Big Data, Educational, Institutions, RapidMiner, Predict Learning, Decision Tree	Evaristus Didik Madyatmadja et al. /Big Data in Educational Institutions using RapidMiner to Predict Learning Effectiveness. Journal of Computer Science 2021, 17 (4): 403.413 DOI: 10.3844/jcssp.2021.	<a href="https://thescipub.com/pdf/jcssp.2021.403.413.pdf">https://thescipub.com/pdf/jcssp.2021.403.413.pdf</a>

			403.413	
4	Big Data and Analytics Implementation in Tertiary Institutions to Predict Students Performance in Nigeria	big data,, analytics, performance, Education,	Big Data And Analytics Implementation In Tertiary Institutions To Predict Students Performance In Nigeria. OC Oguine, KJ Oguine, HI Bisallah - arXiv preprint arXiv:2207.14677, 2022 - arxiv.org	<a href="https://arxiv.org/abs/2207.14677">https://arxiv.org/abs/2207.14677</a>

### Tabla Sociologia Referencia

Sociologia	Título Texto	Palabras Claves	Referência	DOI
1	Six Provocations for Big Data	Big Data, methodology, sociology, computer science, analysis	Boyd, Danah and Crawford, Kate, Six Provocations for Big Data (September 21, 2011). A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society, September 2011, Available at SSRN: <a href="https://ssrn.com/abstract=1926431">https://ssrn.com/abstract=1926431</a>	<a href="http://dx.doi.org/10.2139/ssrn.1926431">http://dx.doi.org/10.2139/ssrn.1926431</a>
2	Citizens' media meets big data: the emergence of data activism	Big data media studies data activism	Milan, S., & Gutiérrez, M. (2015). Citizens' media meets big data: the emergence	<a href="https://doi.org/10.26620/uniminto.mediaciones.11.14.2015.120-133">https://doi.org/10.26620/uniminto.mediaciones.11.14.2015.120-133</a>

			of data activism. Mediaciones, 11(14), 120–133	
3	Le moment big data des sciences sociales	big data Sciences sociales	Bastin, G. & Tubaro, P. (2018) Le moment Big Data des sciences sociales. Revue Française de Sociologie, 59(3), 375-394.	<a href="https://doi.org/10.3917/rfs.593.0375">https://doi.org/10.3917/rfs.593.0375</a>
4	Sobre las limitaciones de Big Data en ciencias sociales	Big Data, redes sociales, técnicas cuantitativas, epistemología, sociología.	MAESTRO CANO, I. (2019). Sobre las limitaciones de Big Data en ciencias sociales. Sociología y tecnociencia.	DOI 10.24197/st.2.2019.77-98
5	La sociología a través de sus publicaciones en revistas de impacto mediante el uso de big data	Sociología, Grandes datos, redes de afiliación, Sociology, big data, Affiliation Networks	Martínez-Uribe, L. (2022). La sociología a través de sus publicaciones en revistas de impacto mediante el uso de big data. Empiria. Revista de metodología de ciencias sociales	<a href="https://doi.org/10.5944/empiria.53.2022.32612">DOI:10.5944/empiria.53.2022.32612</a>

**Tabla Derecho Referencia**

Derecho	Título Texto	Palabras Claves	Referencia	DOI
1	Big Data o la acumulación masiva de datos sanitarios  derechos en riesgo en el marco de la sociedad digital	Big Data, datos personales, protección, derecho a la intimidad	Pérez, M.M. (2015). Big Data o la acumulación masiva de datos sanitarios: derechos en riesgo en el marco de la sociedad digital.	<a href="https://dialnet.unirioja.es/servlet/articulo?codigo=5635389">https://dialnet.unirioja.es/servlet/articulo?codigo=5635389</a>
2	Using big data for legal and law enforcement decisions: Testing the new tools	big data  legal an law	Devins, C., Felin, T., Kauffman, S., & Koppl, R. (2017). The law and big data. Cornell JL & Public Policy, 27, 357	<a href="https://search.informit.org/doi/abs/10.3316/informit.613165001799453">https://search.informit.org/doi/abs/10.3316/informit.613165001799453</a>
3	Derecho y big data	protección de datos  macrodatos  aspectos jurídicos	Becerra, J., Cotino-Hueso, L., León, I. P., Sánchez-Acevedo, M. E., Torres-Ávila, J., & Velandia-Vega, J. (2018). Derecho y big data. Bogotá: Universidad Católica de	<a href="https://hdl.handle.net/10983/22964">https://hdl.handle.net/10983/22964</a>

			Colombia, 2018	
4	Riesgos e impactos del Big Data, la inteligencia artificial y la robótica. enfoques, modelos y principios de la respuesta del derecho.	Big Data, derecho	Hueso, L. C. (2019). Riesgos e impactos del Big Data, la inteligencia artificial y la robótica: enfoques, modelos y principios de la respuesta del derecho. Revista general de Derecho administrativo, (50), 1-37.	<a href="https://dialnet.unirioja.es/servlet/articulo?codigo=6823516">https://dialnet.unirioja.es/servlet/articulo?codigo=6823516</a>
5	LA PROTECCIÓN DE LOS DERECHOS FUNDAMENTALES DE LA PERSONA TRABAJADORA ANTE LA UTILIZACIÓN DE GPS: ¿REFORMULACIÓN O CONTINUIDAD?	GPS, derecho a la protección de datos, derecho a la intimidad, control empresarial, derechos fundamentales.	Garrido, M.E. (2019). La protección de los derechos fundamentales de la persona trabajadora ante la utilización de GPS: ¿reformulación o continuidad? Language, 127-148.	<a href="https://doi.org/10.1387/lan-harremank.21087">DOI:10.1387/lan-harremank.21087</a>
6	Decisiones algorítmicas de perfilado  Régimen y garantías jurídicas	algoritmos  big data  garantía jurídicas	Hernandez, J.C. (2020). Decisiones algorítmicas de perfilado: Régimen y garantías jurídicas.	<a href="https://dialnet.unirioja.es/servlet/articulo?codigo=7479615">https://dialnet.unirioja.es/servlet/articulo?codigo=7479615</a>
7	Análisis documental sobre el tema del big data y su impacto en los derechos humanos	Big data, datos personales, derechos humanos, derecho internacional, privacidad	Téllez Carvajal, Evelyn. (2020). Análisis documental sobre el tema del Big Data y su impacto en los derechos humanos. Derecho PUCP, (84),	<a href="https://dx.doi.org/10.18800/derechopucp.202001.006">https://dx.doi.org/10.18800/derechopucp.202001.006</a>

			155-188.	
--	--	--	----------	--

### Tabla Política Referencia

Política	Título Texto	Palabras Claves	Referência	DOI
1	The Five Vs of Big Data Political Science Introduction to the Virtual Issue on Big Data in Political Science Political Analysis	Virtual Issue, Political Analysis, Big Data	Monroe, B. L. (2013). The five Vs of big data political science introduction to the virtual issue on big data in political science political analysis. Political Analysis, 21(V5), 1-9.	DOI: <a href="https://doi.org/10.1017/S1047198700014315">https://doi.org/10.1017/S1047198700014315</a>
2	Political Campaigns and Big Data	Big Data, Political Campaigns	Nickerson, D. W., & Rogers, T. (2014). Political campaigns and big data. Journal of Economic Perspectives, 28(2), 51-74	DOI: 10.1257/jep.28.2.51
3	Automation, big data and politics: a research review	literature review, critical, big data, algorithms, politics, automation	Shorey, S., & Howard, P. (2016). Automation, big data and politics: A research review. International Journal of Communication, 10.	<a href="https://ora.ox.ac.uk/objects/uuid:252c907d-9357-47bc-b3be-7f2e4ba1fddd">https://ora.ox.ac.uk/objects/uuid:252c907d-9357-47bc-b3be-7f2e4ba1fddd</a>
4	Análisis supervisado de sentimientos	Análisis supervisado de	Arcila-Calderón, C., Ortega-Mohedano,	DOI: <a href="https://doi.org/10.1017/S1047198700014315">https://doi.org/10.1017/S1047198700014315</a>

	políticos en español: clasificación en tiempo real de tweets basada en aprendizaje automático	sentimientos, Opinión política, Twitter, Aprendizaje automático, Big data, Datos masivos, Tweets políticos.	F., Jiménez-Amores, J., & Trullenque, S. (2017). Análisis supervisado de sentimientos políticos en español: clasificación en tiempo real de tweets basada en aprendizaje automático. Profesional de la Información, 26(5), 973-982.	<a href="https://doi.org/10.3145/epi.2017.sep.18">0.3145/epi.2017.sep.18</a>
5	VENTAJAS DE LA UTILIZACIÓN DE BIG DATA EN EL PROCESO AUDITOR, 2018	Big data, Proceso auditor, Entidades de fiscalización superiores	López Espinosa, J.N. (2022). VENTAJAS DE LA UTILIZACIÓN DE BIG DATA EN EL PROCESO AUDITOR, 2018. REVISTA CHILENA DE LA ADMINISTRACIÓN DEL ESTADO.	<a href="https://doi.org/10.57211/revista.v1i01.34">DOI:10.57211/revista.v1i01.34</a>

### Tabla Geografía Referencia

Geografía	Título Texto	Palabras Claves	Referencia	DOI
1	Geography and the future of big data, big data and the future of geography	big data geography	Graham, M., & Shelton, T. (2013). Geography and the future of big data, big data and the future of geography. Dialogues in Human geography, 3(3), 255-261.	<a href="https://doi.org/10.1177/2043820613513121">https://doi.org/10.1177/2043820613513121</a>

2	Big data and human geography: Opportunities, challenges and risks	big data risk geography challenges	Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. <i>Dialogues in human geography</i> , 3(3), 262-267	<a href="https://doi.org/10.1177/2043820613513388">https://doi.org/10.1177/2043820613513388</a>
3	Considerations on Geospatial Big Data	geospatial big data	Zhen LIU et al 2016 IOP Conf. Ser.: Earth Environ. Sci. 46 012058	DOI 10.1088/1755-1315/46/1/012058
4	Big Data y nuevas geografías: la huella digital de las actividades humanas	Big Data et nouvelles géographies : l'empreinte digitale des activités humaines	Puebla, J. G. (2018). Big Data y nuevas geografías: la huella digital de las actividades humanas. <i>Documents d'anàlisi geogràfica</i> , 64(2), 195-217	<a href="https://doi.org/10.5565/rev/dag.526">https://doi.org/10.5565/rev/dag.526</a>
5	Big Data and Geospatial Analysis	big data geo analysis	Lansley, G., de Smith, M., Goodchild, M., & Longley, P. (2019). Big data and geospatial analysis. arXiv preprint arXiv:1902.06672.	<a href="https://doi.org/10.48550/arXiv.1902.06672">https://doi.org/10.48550/arXiv.1902.06672</a>

### Tabla Periodismo Referencia

Periodismo	Título Texto	Palabras Claves	Referència	DOI
1	The use of big data in corporate communication	Public Relations Marketing Communications Corporate	Wiencierz, C. and Röttger, U. (2017), "The use of big data in corporate communication", Corporate	<a href="https://doi.org/10.1108/CCIJ-02-2016-0015">https://doi.org/10.1108/CCIJ-02-2016-0015</a> <a href="#">Download as RIS</a>

		communications Big data Systematic review	Communications: An International Journal, Vol. 22 No. 3, pp. 258-272.	
2	Big Data Analytics for Emergency Communication Networks: A Survey	Big data emergency network	Wang, J., Wu, Y., Yen, N., Guo, S., & Cheng, Z. (2016). Big data analytics for emergency communication networks: A survey. IEEE Communications Surveys & Tutorials, 18(3), 1758-1778	<b>DOI:</b> <a href="https://doi.org/10.1109/COMST.2016.2540004">10.1109/COMS T.2016.2540004</a>
3	Algoritmos, aplicaciones y Big data, nuevos paradigmas en el proceso de comunicación y de enseñanza-aprendizaje del periodismo de datos	periodismo; datos; investigación; aplicaciones; algoritmos.	Vivar, J.M. (2018). Algoritmos, aplicaciones y Big data, nuevos paradigmas en el proceso de comunicación y de enseñanza-aprendi zaje del periodismo de datos.	<a href="https://doi.org/10.26441/RC17.2-2018-A12">DOI:10.26441/ RC17.2-2018-A 12</a>
4	‘Issues’ y ‘big data’ en la gestión de relaciones públicas. El caso de la implementación del nuevo sistema de tratamiento de residuos “Recuperando Valor” en Córdoba, Argentina	gestión big data public	Castillo-Esparcia, A., Álvarez-Nobell, A., & Barroso, M.B. (2021). ‘Issues’ y ‘big data’ en la gestión de relaciones públicas. El caso de la implementación del nuevo sistema de tratamiento de residuos “Recuperando Valor” en Córdoba, Argentina.	DOI:10.51698/ TRIPODOS.20 19.45P73-87

			Tripodos,	
--	--	--	-----------	--

**Tabla Asociación término big data en cada textos Cs + AL + 2020**

texto	ano	n° x big data	pq hablan de big data?	teórico x metodológico
26	2014	5 vez es	<p><b>big data 0:</b> abstract</p> <p><b>big data 1:</b> grande cantidad d edatos (enormes conjuntos de datos, como la bioinformática, la quimioinformática y la neuroinformática): determinando nueva division del tabajo en la ciencia(os recopiladores de datos y los analistas de datos (los científicos se esfuerzan por conseguir algo parecido)</p> <p><b>big data 2:</b> nuevo modleo de producion de datos científicos (e han dejado llevar por la explosión de datos, imágenes y virtualización en la investigación científica hasta el punto de afirmar, quizás no del todo como boutades, que la era de la teoría en las ciencias ha terminado con la instauración de los big data)</p> <p><b>big data 3:</b> revolucion informatica ligada a “explosion” de los grandes volumnes de datos</p> <p><b>big data 4:</b> extito predictivos en corto plazo, que aun esta por llegar con los big data</p>	<b>teórico</b>
25	2016	23	23 palavras: texto explicativo : presenta el fenomeno. Levanta conceptos al rededro del termino y problematiza el fututo de la sociologia a partir de la apropiacion de la digitalidad d elos datos y capacidad computacional	teorico

24	2017	5	gigantescos sistemas de absorcion de datos; fin de la privacidad; permite hacer pronostico de comportamiento humano; efectos na produccion de transparencia; capitalismo de la transparencia  es un articulo preocupado con los riesgos de la implementacion de big data y apunta a sus cuidados y precauciones.	teorico
23	2017	8	una emergente realidad; medias sociales y informacion personal; picos de busca en internet; mas que grandes datos, un conjunto de pensamiento; cambiando interacciones humanas	metodológico
22	2017	9	comparacion con small data; huellas digitales y sobreinformacion; dispositivos de mercado; produccion de informacion social cuantitativa: re.semblar la informacion social; practicas financieras; buen habitos en relacion a registro de informacion; desafio con relacion a naturaleza del conocimiento social	teorico
21	2017	10	investigacion en educacion; big data educacionales; tecnica emergente ; e-learning; conversion de informacion; ventajas procesamiento transacciones electronicas;	metodologico
20	2017	6	Interpretacion de grandes cantidades de datos; trabajos del futuro; produccion y mantenimiento de grandes cantidades de datos	teórico
19	2017	18	ensenanza aprendizaje musical; sistema de gestion de conocimientos; tecnicas big data; modelos para grandes datos; plataforma telemática; grandes masas de datos; limpieza y seleccion de informacion; monitoreo de aprendizaje musical; analisis de datos educativos; cursos masivos en linea	metodológico
18	2018	7	almacenamiento grandes cantidades de datos; avance tecnologico; accion de internautas; psicometria; manipulacion de la red; psicoprograma colectivo	teórico
17	2018	2	expansión de la ciencia y inteligencia artificial	teórico

16	2018	26	algoritmos matemáticos; elecciones anti democráticas; determinación de contenido; mercado de la política; datos masivos; capacidad de aprovechar información; volumen, velocidad, variedad de datos; explicación de acontecimiento y no de los porqués; presiones monopolistas; intensificación de ventas y explotación de mercados; riesgos humanidades digitales; reconocimiento de patrones; Ciência Social Quantitativa da Escola de Negócios; control de capital; opiniones transformadas en matemática; radicalizar explotación del usuario;	teórico
15	2019	11	sistema educacional ahogado en información; nuevas iniciativas de recopilación de datos; tendencia en política educativa; minimizan la discusión de las dificultades metodológicas de la evaluación de impacto; no han reemplazado el elemento humano en la toma de decisión; oculta sesgos en el aspecto político; desafía estatus quo de la gobernanza global	teórico
14	2019	19	relación entre política y redes sociales; revista big data & society; minería de datos de comportamientos humanos; emergencia como imperativo en la ciencia; encuesta de opinión; opera base de datos de otros propósitos; lógica de la información; agenda de investigación; generación continua de conocimiento	teórico
13	2019	4	datificación; mundo de los datos, predominancia de datos	teórico
12	2019	2	mineración de datos, uso masivo de información	teórico
11	2019	2	polarización del debate políticos; técnicas big data	
10	2019	2	debate sobre los métodos e técnicas de pesquisa empírica nas ciências sociais; datificación	teórico
09	2020	3	características positivas na construção de novos indicadores; ferramentas de data analítica; medir grado de consumo	teórico
08	2020	19	datos y computación en nube; internet de las cosas; inteligencia artificial; big data analytics; marketing,	teórico

			inteligência y negocios; algoritmos de aprendizaje de maquina; datos unipresentes; crecimiento de volume d edtaos; computacion en nube	
7	2020	1	survey big data para retrato nacional de psicoanálisis.	teorico
8	2020	10	vigilancia capitalismo; emergencia de grandes datos; afecta y redefine la sociedad contemporanea; acumulacion de capital en forma de datos; monopolio de grandes empresas multinacionles como google; poder predictivo;  associada a las redes sociales y la capura de informcaion personal; relacion de big data con la agenda de vigilancia, capitalista.	teorico
6	2020	32	tecnica viable que permite brindar respuesta a la problematica de desercion universitaria; modelo analaitico big data; características del <i>Big Data</i> para poder identificar y potencializar sus fun- cionalidades ; volumen, variabilidad y velocidad. Por otra parte, las “V” relacionadas con veracidad, valor y visualización; soluciones tecnologicas; puede trabajar con datos estructurados y no estructurado; prevalencia d edato estructurado; modleo analitico; cientifico de datos como parte del equipo big data; solucion tecnologica; herramienta de minerai de datos; producción de alertas y advertencias de notificacion ; estrategia viable y pertinente para revertir el problema de la deserción escolar; gran importancia en la educación; posibilidades de elaborar estrategias educativas con base en datos para mejorar y garantizar la efectividad del aprendizaje	metodologico
5 6	2020	23	enriquece la comprensión sobre los fenómenos globales; nueva promessa metodologica; desaio metodologico para los estudios globales; tecnologías disruptivas que han modificado considerables aspectos de la sociedad; facilitado los avances recientes en la medicina de precisión; estudio de la comunicación verbal y no verbal, tanto con métodos computacionales como cognitivos; procesamiento de datos; Cox y Ellsworth (1997, 5-7), quienes lo definieron como el proceso de graficar, para su correcta visualización, grandes cantidades de información, tiene valor por seren inteligentes; un fenómeno cultural, tecnológico y académico que se basa en la interacción de la	teorico

			tecnología, el análisis y mitología; reto par ainvestigaciones sociales; minería de datos; campo d einvestigacion completamente diferente;  potente herramienta para el análisis de las tendencias modernas.	
4	2020	5	desarrollo de sistemas integra- dos de información; análisis tecnológicas, inteligencias artificiales; sistemas integrados; herramientas tecnológicas que fusionan la informa- ción y las telecomunicaciones; exito en big data a nivel policial; sistema integrado de fuentes d einformacion; Datapol	metodologico
3	2020	7	fuelle de datos con diferentes características; planea- ción urbana y políticas públicas; datos de movilidad: funcionalidades con uso intelligen- cia artificial, análisis de datos; modelaciones semafóricas;	metodologico
2	2020	6	poderosa tecnología algorítmica movida a enormes reservas de datos; interpretar básicamente dois fenômenos culturais contemporâneos; automatizacion d eprocessos, acesibilidad integral a todos; atividade economicas; totalitarismo de mercado; revolucion publicitária;	teorico
1	2020	2	dataismo; grandes datos	teorico

### Los 5 principales contribuyentes por región y país

Región	Autor	País	Publicaciones
USA	Ye X.	USA	15
USA	Li Z.	USA	15
USA	Gao J.	USA	11
USA	Yang C.	USA	11
USA	Li W.	USA	11
RUS	Berezkin Y.	RUS	6
RUS	Zhuravleva N.	RUS	4
RUS	Kotenko I.	RUS	3
RUS	Kupriyanov A.	RUS	3
RUS	Popkova E.	RUS	3
OCEANIA	Pettit C.	AUS	15
OCEANIA	Sinnott R.	AUS	8
OCEANIA	Daniel B.	NZL	7
OCEANIA	Xia F.	AUS	7
OCEANIA	Strang K.	AUS	7
EUROPA	Jiang B.	SWE	14
EUROPA	Mosavi A.	DEU	13
EUROPA	Taylor L.	NLD	12
EUROPA	Poorthuis A.	BEL	11
EUROPA	Zwitter A.	NLD	10
EUR NO-UE	Bibri S.	NOR	24
EUR NO-UE	Batty M.	GBR	19
EUR NO-UE	Williamson B.	GBR	12

EUR NO-UE	Krogstie J.	NOR	10
EUR NO-UE	Schroeder R.	GBR	9
CHN	Li Q.	CHN	20
CHN	Liu Y.	CHN	19
CHN	Tu W.	CHN	11
CHN	Niu L.	CHN	11
CHN	Pei T.	CHN	10
CAN	Boulanger D.	CAN	8
CAN	Seanosky J.	CAN	7
CAN	Chow-White P.	CAN	6
CAN	Gaboury S.	CAN	6
CAN	Winne P.	CAN	5
ASIA	Zhou J.	HKG	10
ASIA	Kwan M.	HKG	8
ASIA	Law R.	HKG	7
ASIA	Lu Y.	HKG	7
ASIA	Yeh A.	HKG	6
ASIA	Anshari M.	BRN	9
ASIA	Shibasaki R.	JPN	8
ASIA	Park H.	KOR	8
ASIA	Ratti C.	SGP	8
ASIA	Gani A.	MYS	7
ASIA	Lytras M.	SAU	7

ASIA	Hasan R.	PAK	6
ASIA	Aljohani N.	SAU	5
ASIA	Hashem I.	ARE	5
ASIA	Mehmood R.	PAK	4
ALC	Maçada A.	BRA	5
ALC	Câmara G.	BRA	5
ALC	Vinhas L.	BRA	4
ALC	Hoppen N.	BRA	3
ALC	Cacho N.	BRA	3
ALC	Munizaga M.	CHL	4
ALC	Villegas-Ch W.	ECU	4
ALC	Cervantes O.	MEX	3
ALC	Lu K.	PRI	3
ALC	Ochoa D.	ECU	2
AFRICA	Hassanien A.	EGY	5
AFRICA	Mutula S.	ZAF	4
AFRICA	Ajibade P.	ZAF	4
AFRICA	Prinsloo P.	ZAF	3
AFRICA	Boudriga N.	TUN	3

Las 5

**principales instituciones que investigan sobre Big Data por región**

<b>Region</b>	<b>Institución</b>	<b>País</b>	<b>Publicaciones</b>
USA	Arizona State University	USA	24

USA	University of California, Santa Barbara	USA	23
USA	Texas A&M University	USA	22
USA	University of Illinois Urbana-Champaign	USA	18
USA	University of South Carolina, Department of Geography	USA	18
RUS	HSE University	RUS	13
RUS	Saint Petersburg State University	RUS	8
RUS	Financial University under the Government of the Russian Federation	RUS	8
RUS	Lomonosov Moscow State University	RUS	7
RUS	The State University of Management	RUS	7
OCEANIA	UNSW Sydney	AUS	33
OCEANIA	Queensland University of Technology	AUS	17
OCEANIA	University of Technology Sydney	AUS	17
OCEANIA	The University of Sydney	AUS	17
OCEANIA	RMIT University	AUS	14
Union Europea	Technical University of Munich	DEU	24
Union Europea	KU Leuven	BEL	22
Union Europea	University of Žilina, Department of Economics	SVK	22
Union Europea	Universiteit van Amsterdam	NLD	18
Union Europea	Technische Universität Dresden	DEU	18
EUR NO-UE	Cardiff University	GBR	20
EUR NO-UE	University College London	GBR	32

EUR NO-UE	Norges Teknisk-Naturvitenskapelige Universitet, Department of Architecture and Planning	NOR	24
EUR NO-UE	Norges Teknisk-Naturvitenskapelige Universitet	NOR	17
EUR NO-UE	Universität Zürich	CHE	11
China	University of Chinese Academy of Sciences	CHN	65
China	Peking University	CHN	56
China	Tsinghua University	CHN	46
China	Wuhan University	CHN	44
China	Beihang University	CHN	36
Canada	University of Toronto	CAN	18
Canada	University of Manitoba	CAN	17
Canada	Université McGill	CAN	12
Canada	University of Calgary	CAN	9
Canada	Queen's University	CAN	9
Hong Kong	Hong Kong Polytechnic University	HKG	42
Hong Kong	City University of Hong Kong	HKG	20
Hong Kong	The University of Hong Kong, Department of Urban Planning and Design	HKG	20
Hong Kong	The University of Hong Kong	HKG	13
Hong Kong	School of Hotel and Tourism Management, The Hong Kong Polytechnic University	HKG	16
ASIA	The University of Tokyo	JPN	16

ASIA	Vellore Institute of Technology	IND	14
ASIA	National University of Singapore	SGP	13
ASIA	Universiti Malaya	MYS	12
ASIA	Nanyang Technological University	SGP	10
Medio Oriente	Tel Aviv University	ISR	10
Medio Oriente	King Abdulaziz University	SAU	9
Medio Oriente	Middle East College, Department of Computing	OMN	8
Medio Oriente	King Abdulaziz University, Faculty of Computing and Information Technology	SAU	7
Medio Oriente	Sir Syed University of Engineering and Technology	PAK	6
Brasil	Universidade de São Paulo	BRA	20
Brasil	Universidade Federal do Rio Grande do Sul	BRA	7
Brasil	Universidade Estadual Paulista "Júlio de Mesquita Filho"	BRA	6
Brasil	Instituto Nacional de Pesquisas Espaciais	BRA	6
Brasil	Universidade Federal de Santa Catarina	BRA	6
America Latina	Universidad Nacional Autónoma de México	MEX	6
America Latina	Tecnologico de Monterrey	MEX	4
America Latina	Pontificia Universidad Católica de Chile	CHL	4

America Latina	Pontificia Universidad Catolica del Peru	PER	4
America Latina	Universidad de las Americas - Ecuador	ECU	4
AFRICA	University of Johannesburg	ZAF	8
AFRICA	University of South Africa	ZAF	6
AFRICA	Cairo University	EGY	6
AFRICA	University of KwaZulu-Natal, Information Studies Programme	ZAF	5
AFRICA	Faculté des Sciences Rabat	MAR	4

**Principales revistas científicas que publican artículos relacionados con "Big Data" en ciencias sociales por regiones.**

Región	Journal	Publicaciones
USA	Big Data and Society	44
USA	Computers, Environment and Urban Systems	25
USA	ISPRS International Journal of Geo-Information	21
USA	Sustainability (Switzerland)	19
USA	Transportation Research Part C: Emerging Technologies	18
RUS	Monitoring Obshchestvennogo Mneniya: Ekonomicheskie i Sotsial'nye Peremeny	17
RUS	Sotsiologicheskie Issledovaniya	6
RUS	Antropologicheskij Forum	3
RUS	Computer Law and Security Review	3
RUS	Finance: Theory and Practice	3
OCEANIA	Big Data and Society	11
OCEANIA	Geopolitics, History, and International Relations	5

OCEANIA	Journal of Computer Information Systems	5
OCEANIA	Journal of Self-Governance and Management Economics	5
OCEANIA	Sustainability (Switzerland)	5
EUROPA	Sustainability (Switzerland)	95
EUROPA	Big Data and Society	87
EUROPA	ISPRS International Journal of Geo-Information	35
EUROPA	International Journal of Information Management	32
EUROPA	Profesional de la Informacion	29
CHN	Sustainability (Switzerland)	112
CHN	ISPRS International Journal of Geo-Information	81
CHN	International Journal of Emerging Technologies in Learning	35
CHN	Jiaotong Yunshu Xitong Gongcheng Yu Xinxi/Journal of Transportation Systems Engineering and Information Technology	35
CHN	Dili Xuebao/Acta Geographica Sinica	33
ASIA	Sustainability (Switzerland)	159
ASIA	International Journal of Scientific and Technology Research	71
ASIA	Advanced Science Letters	40
ASIA	ISPRS International Journal of Geo-Information	21
ASIA	Sustainable Cities and Society	17
ALC	Perspectivas em Ciencia da Informacao	6
ALC	Ciencia da Informacao	5
ALC	Sustainability (Switzerland)	5
ALC	Research in Transportation Economics	3
ALC	Revista Brasileira de Politicas Publicas	3
AFRICA	Library Philosophy and Practice	4

AFRICA	International Journal of Emerging Technologies in Learning	3
AFRICA	International Review of Research in Open and Distance Learning	3
AFRICA	Potchefstroom Electronic Law Journal	3
AFRICA	African Journal of Hospitality, Tourism and Leisure	2

**Principales conferencias científicas en las que se han presentado trabajos relacionados con los "Big data" en ciencias sociales por regiones**

Región	Conferencia	Publicaciones
USA	2017 IEEE SmartWorld	88
USA	2016 IEEE SustainCom	34
USA	2015 IEEE SC2	19
USA	5th IEEE BigDataService 2019	18
USA	2020 IEEE ISPA-BDCloud-SocialCom-SustainCom 2020	15
RUS	Komp'juternaja Lingvistika i Intellektual'nye Tehnologii	3
RUS	2015 IEEE SC2 2015	3
RUS	2019 IEEE ComSDS	3
RUS	2017 IEEE SmartWorld	2
RUS	2019 4th ICBDS	2
OCEANIA	2015 IEEE SC2 2015	11
OCEANIA	Lecture Notes in Geoinformation and Cartography	7
OCEANIA	BDCAT 2019	5

OCEA NIA	ISPRS Archives	4
OCEA NIA	13th IEEE eScience 2017	4
EURO PA	2017 IEEE SmartWorld	65
EURO PA	ISPRS Archives	51
EURO PA	2015 IEEE SC2	20
EURO PA	Transportation Research Procedia	18
EURO PA	Lecture Notes in Geoinformation and Cartography	17
CHN	2015 ICITBS	246
CHN	2020 ICITBS	221
CHN	2019 IEEE ISPA/BDCloud/SustainCom/SocialCom	195
CHN	3rd ICITBS 2018	188
CHN	2019 ICITBS	160
ASIA	2015 ICSITech	63
ASIA	2015 IEEE SC2 2015	53
ASIA	2016 ICBDS	52
ASIA	2017 IEEE SmartWorld	45
ASIA	2019 4th ICBDS	36
ALC	ISPRS Archives	7
ALC	2017 IEEE SmartWorld	4
ALC	Proceedings of the Brazilian Symposium on GeoInformatics	4
ALC	Proceedings of the LACCEI international Multi-conference for Engineering, Education and Technology	4

ALC	2020 CONIITI	2
AFRIC A	IBASE-BF 2020 - 1st IEEE	9
AFRIC A	ISPRS Archives	7
AFRIC A	2021 IST-Africa Conference	5
AFRIC A	2020 IST-Africa Conference	4
AFRIC A	2017 IEEE SmartWorld	3

**Cuadro X: Los 5 autores de ciencias sociales más citados sobre "Big Data" por región**

<b>Autor</b>	<b>Región</b>	<b>País</b>	<b>Afiliación</b>	<b>Cited by</b>
Crawford K.	USA	USA	University of Southern California	517
Zook M.	USA	USA	University of Kentucky	492
Seligman M.	USA	USA	University of Pennsylvania	357
Dhar V.	USA	USA	Leonard N. Stern School of Business	349
Hilbert M.	USA	USA	University of California, Davis	325
Soldatova G.	RUS	RUS	Lomonosov Moscow State University	39
Zhukov D.	RUS	RUS	MIREA - Russian Technological University (RTU MIREA)	17
Anikin I.	RUS	RUS	Kazan National Research Technical University named after A. N. Tupolev -KAI	16
Suyatinov S.	RUS	RUS	Bauman Moscow State Technical University	12

Nikolenko S.	RUS	RUS	Saint Petersburg State University	12
Andrejevic M.	OCEANIA	AUS	Monash University	616
Daniel B.	OCEANIA	NZL	University of Otago	388
Allam Z.	OCEANIA	AUS	Deakin University	258
Lingard B.	OCEANIA	AUS	Australian Catholic University	224
Stantić B.	OCEANIA	AUS	Griffith University	174
Boyd D.	EUROPA	GBR	Microsoft Research Cambridge	2879
Kitchin R.	EUROPA	IRL	Maynooth University	2553
Batty M.	EUROPA	GBR	University College London	1118
Floridi L.	EUROPA	GBR	Alan Turing Institute	757
van Dijck J.	EUROPA	NLD	Universiteit Utrecht	609
De Mauro A.	EUROPA	ITA	Università degli Studi di Roma "Tor Vergata"	506
Kang C.	CHN	CHN	Wuhan University	527
Liu X.	CHN	CHN	Sun Yat-Sen University	424
Pei T.	CHN	CHN	Institute of Geographical Sciences and Natural Resources Research Chinese Academy of Sciences	347
Wang S.	CHN	CHN	Chinese Academy of Sciences	296
Ran B.	CHN	CHN	Southeast University	282
Anuar N.	ASIA	MYS	Universiti Malaya	778
Chau K.	ASIA	HKG	Hong Kong Polytechnic University	374
Ogata H.	ASIA	JPN	Kyoto University	274
Law R.	ASIA	HKG	School of Hotel and Tourism Management, The Hong Kong Polytechnic University	217
Meier P.	ASIA	QAT	Qatar Computing Research Institute	211

Evsukoff A.	ALC	BRA	Universidade Federal do Rio de Janeiro	246
Bonilla S.	ALC	BRA	Universidade Paulista	183
Tavares E.	ALC	BRA	Universidade Federal do Rio de Janeiro	167
Estrella J.	ALC	BRA	Universidade de São Paulo	62
Telles R.	ALC	BRA	Universidade Paulista	54
Munizaga M.	ALC	CHL	Instituto Sistemas Complejos de Ingeniería	53
Ríos S.	ALC	CHL	Universidad de Chile	34
Cervantes O.	ALC	MEX	Universidad de las Américas Puebla	29
Forcael E.	ALC	CHL	Universidad del Bio Bio	27
Ochoa-Zezzati A.	ALC	MEX	Universidad Autónoma de Ciudad Juárez	26
Prinsloo P.	AFRICA	ZAF	University of South Africa	56
Alimi A.	AFRICA	TUN	University of Sfax	43
Badir H.	AFRICA	MAR	Ecole Nationale des Sciences Appliquées de Tanger	38
Mouline S.	AFRICA	MAR	Faculté des Sciences Rabat	37
Mashwama P.	AFRICA	SWZ	University of Eswatini	30

Los 5 autores de ciencias sociales más citados sobre "Big Data" por región

<b>Autor</b>	<b>Región</b>	<b>País</b>	<b>Afiliación</b>	<b>Cited by</b>
Crawford K.	USA	USA	University of Southern California	517
Zook M.	USA	USA	University of Kentucky	492
Seligman M.	USA	USA	University of Pennsylvania	357

Dhar V.	USA	USA	Leonard N. Stern School of Business	349
Hilbert M.	USA	USA	University of California, Davis	325
Soldatova G.	RUS	RUS	Lomonosov Moscow State University	39
Zhukov D.	RUS	RUS	MIREA - Russian Technological University (RTU MIREA)	17
Anikin I.	RUS	RUS	Kazan National Research Technical University named after A. N. Tupolev -KAI	16
Suyatinov S.	RUS	RUS	Bauman Moscow State Technical University	12
Nikolenko S.	RUS	RUS	Saint Petersburg State University	12
Andrejevic M.	OCEANIA	AUS	Monash University	616
Daniel B.	OCEANIA	NZL	University of Otago	388
Allam Z.	OCEANIA	AUS	Deakin University	258
Lingard B.	OCEANIA	AUS	Australian Catholic University	224
Stantić B.	OCEANIA	AUS	Griffith University	174
Boyd D.	EUROPA	GBR	Microsoft Research Cambridge	2879
Kitchin R.	EUROPA	IRL	Maynooth University	2553
Batty M.	EUROPA	GBR	University College London	1118
Floridi L.	EUROPA	GBR	Alan Turing Institute	757
van Dijck J.	EUROPA	NLD	Universiteit Utrecht	609
De Mauro A.	EUROPA	ITA	Università degli Studi di Roma "Tor Vergata"	506
Kang C.	CHN	CHN	Wuhan University	527
Liu X.	CHN	CHN	Sun Yat-Sen University	424
Pei T.	CHN	CHN	Institute of Geographical Sciences and Natural Resources Research Chinese Academy of Sciences	347

Wang S.	CHN	CHN	Chinese Academy of Sciences	296
Ran B.	CHN	CHN	Southeast University	282
Anuar N.	ASIA	MYS	Universiti Malaya	778
Chau K.	ASIA	HKG	Hong Kong Polytechnic University	374
Ogata H.	ASIA	JPN	Kyoto University	274
Law R.	ASIA	HKG	School of Hotel and Tourism Management, The Hong Kong Polytechnic University	217
Meier P.	ASIA	QAT	Qatar Computing Research Institute	211
Evsukoff A.	ALC	BRA	Universidade Federal do Rio de Janeiro	246
Bonilla S.	ALC	BRA	Universidade Paulista	183
Tavares E.	ALC	BRA	Universidade Federal do Rio de Janeiro	167
Estrella J.	ALC	BRA	Universidade de São Paulo	62
Telles R.	ALC	BRA	Universidade Paulista	54
Munizaga M.	ALC	CHL	Instituto Sistemas Complejos de Ingeniería	53
Ríos S.	ALC	CHL	Universidad de Chile	34
Cervantes O.	ALC	MEX	Universidad de las Américas Puebla	29
Forcael E.	ALC	CHL	Universidad del Bio Bio	27
Ochoa-Zezzatti A.	ALC	MEX	Universidad Autónoma de Ciudad Juárez	26
Prinsloo P.	AFRICA	ZAF	University of South Africa	56
Alimi A.	AFRICA	TUN	University of Sfax	43
Badir H.	AFRICA	MAR	Ecole Nationale des Sciences Appliquées de Tanger	38
Mouline S.	AFRICA	MAR	Faculté des Sciences Rabat	37
Mashwama P.	AFRICA	SWZ	University of Eswatini	30

### Guión- Entrevistas con autores

Preguntas Generales:

- 1- Trayectoria Académica
- 2- Vínculos institucionales y laborales
- 3- Interés por el tema big data
- 4- Es usuario de tecnologías big data/sabe programar
- 5- Cómo llegó a los datos que presentó
- 6- Dificultades de trabajar con el tema
- 7- Oportunidades de trabajo alrededor de big data
- 8- Porque escogió publicar la parte específica que habla de big data, aún sabiendo que es una parte muy pequeña de la investigación.
- 9 - Hablar sobre: proceso de adaptación a las metodologías y herramientas
- 10- Cómo convivir en un mundo entre las ciencias sociales y tecnológicas: problema de Traducción
- 11- ¿Sigues trabajando/estudiando la temática big data?
- 12- sí sí, sientes que desde la fecha de la escritura de la publicación, hasta la actualidad, ¿qué cosas percibes que han cambiado en este universo?
- 13- Necesitaste de auxilio en alguna parte de la producción del artículo y generación de datos (para caso de artículos metodológicos)?