



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

ESTUDIO SOBRE EL ACCESO ABIERTO A DATOS GUBERNAMENTALES: EL
CASO DE TRANSPARENCIA EN CHILE, COSTA RICA Y URUGUAY

TESIS PARA OPTAR AL GRADO DE
MAGÍSTER EN CIENCIAS, MENCIÓN COMPUTACIÓN

MARCOS ALEXIS VENEGAS ÁLVAREZ

PROFESOR GUÍA:
CLAUDIO GUTIÉRREZ GALLARDO

MIEMBROS DE LA COMISIÓN:
PABLO GONZÁLEZ JURE
JOSÉ PINO URTUBIA
ELIANA SCHEIHING GARCÍA

SANTIAGO DE CHILE
2018

Resumen

En América Latina y alrededor del mundo las iniciativas de transparencia y gobierno abierto han aumentando en los últimos años. Los gobiernos están publicando datos abiertos para transparentar su administración, apoyar la toma de decisiones y promover el desarrollo. Es por ello, que los datos abiertos se han convertido en objeto de análisis y estudio.

Esta tesis aborda el estudio de la apertura, capacidad de reutilización y accesibilidad de datos abiertos en algunos gobiernos de América Latina.

En la primera parte de este trabajo se desarrolla el tema de datos, información y conocimiento, enfatizando las diferencias entre datos e información. También se profundiza en el tema de transparencia y su relación con la corrupción, el acceso a la información pública y el gobierno abierto.

Posteriormente, se realiza una revisión bibliográfica sobre datos abiertos y gobierno abierto. También se estudian los usuarios de datos abiertos. Además, se plantean los posibles tipos de usos que los usuarios les pueden dar a los datos abiertos. Se hace una revisión de la literatura sobre dimensiones de datos (en sistemas de información cerrados) y a partir de ellas se propone dimensiones para los datos abiertos.

Por otro lado, se realiza una revisión de la literatura sobre las métricas de evaluación de portales e información de transparencia. Considerando las dimensiones de datos abiertos planteadas, se propone una metodología genérica y métricas que permitan evaluar el nivel de apertura, usabilidad y accesibilidad de los datos de gobierno, particularmente los de transparencia.

Por último, se valida la metodología evaluando el nivel de apertura, capacidad de reutilización y accesibilidad de los conjuntos de datos abiertos publicados en los sitios de transparencia de los gobiernos de Chile, Costa Rica y Uruguay. En la validación también se evalúan algunos sitios de estadística de los gobiernos que poseían datos abiertos. Esta validación concluye con un informe preliminar sobre datos abiertos de esos gobiernos, y con algunas recomendaciones sobre la publicación datos abiertos por parte de los gobiernos.

A Dios, porque su palabra dice: “Y todo lo que hagáis, hacedlo de corazón, como para el Señor y no para los hombres”. Y yo agrego con excelencia para Dios y para los hombres.

A mi esposa María, a quien amo y por quien día a día me esfuerzo. A mis futuros hijos, quienes por este estudio se les ha retrasado su venida a este mundo.

A mi hermano Jose, mi hermana Yahaira y mi sobrino Diberth. A quienes también amo profundamente y deseo serle un buen ejemplo.

A los investigadores de este tema, y a quienes publican y usan datos abiertos. Espero haberles contribuido con mi trabajo “este granito de arena”.

Agradecimientos

A Dios, por darme vida, salud y la oportunidad de estar en este lugar para estudiar e investigar. Sin su bendición hubiera sido imposible estar aquí y concluir mi trabajo.

A mi esposa María López, por su paciencia y sabiduría, por apoyarme en mis sueños y metas.

A la *Universidad de Costa Rica*, por el apoyo financiero brindado para mi formación académica en este país. A los compañeros de la Oficina de Asuntos Internacionales y Cooperación Externa de la Universidad de Costa Rica, por su diligente trabajo, su comprensión y apoyo brindado. A mis compañeros de la Asamblea de la Sede del Pacífico de la Universidad de Costa Rica, por el apoyo y confianza que me brindaron.

A Eduardo Chinchilla y Gabriel Rueda, por su amistad, confianza y apoyo. Sin su ayuda no hubiera sido posible estar aquí.

A mi hermano Jose Venegas, por ayudarme en todas las gestiones en Costa Rica, mientras yo he estado acá estudiando.

A mi tutor Claudio Gutiérrez, por su apoyo sincero, por trabajar con excelencia y brindarme la oportunidad de investigar a su lado.

A todos los hombres y mujeres que han investigado sobre datos, datos abiertos, transparencia y gobierno abierto. En especial, a quienes han puesto su trabajo de forma gratuita para que otros continuemos investigando.

A Freepik <http://www.freepik.com>, por otorgar licencia libre a algunos de sus trabajos, los cuales utilicé para construir esta tesis.

Tabla de Contenido

1. Introducción	1
1.1. Objetivo general	2
1.2. Objetivos específicos	3
1.3. Plan y metodología	3
1.4. Estructura del presente documento	3
2. Datos abiertos, Gobierno Abierto y Transparencia	5
2.1. Datos	5
2.1.1. Datos, información y conocimiento	6
2.1.2. Productores/Usuarios/Intermediarios de datos o Datamediarios	7
2.1.3. Ecosistemas de datos	8
2.2. Datos abiertos	9
2.2.1. La tendencia “Open” y los datos abiertos	9
2.2.2. Datos enlazados	14
2.2.3. Gobierno Abierto	16
2.3. Transparencia	18
2.3.1. Transparencia activa y pasiva	20
2.3.2. Transparencia y acceso a la información	21
2.3.3. Transparencia y corrupción	21
2.3.4. Principales hitos de transparencia en Chile, Costa Rica y Uruguay	23
2.3.5. Trabajos relacionados sobre evaluación de la transparencia	29
3. Producción, usos, usuarios y dimensiones de los datos abiertos	31
3.1. Usos y usuarios de los datos abiertos	31
3.1.1. Producción de datos de forma estática y dinámica	33
3.1.2. Usuarios técnicos y no técnicos	35
3.1.3. Relaciones entre usos y usuarios	35
3.2. Dimensiones de datos	36
3.3. Dimensiones de calidad de datos de un portal Web	43
3.4. Dimensiones de datos abiertos	45
3.5. Dimensiones de datos abiertos según usos y usuarios	47
3.5.1. Descripción de las dimensiones	59
4. Métricas e indicadores para datos abiertos	62
4.1. Revisión de métricas de información de transparencia	62
4.2. Indicadores para la evaluación de datos abiertos según dimensiones estudiadas	69

4.3. Dimensiones no evaluadas	83
5. Resultados y análisis	84
5.1. Identificación de las fuentes de datos	84
5.1.1. Fuentes de datos en educación	86
5.1.1.1. Educación en Uruguay	86
5.1.1.2. Educación en Costa Rica	91
5.1.1.3. Educación en Chile	92
5.1.2. Fuentes de datos en Salud	93
5.1.2.1. Salud en Uruguay	93
5.1.2.2. Salud en Costa Rica	93
5.1.2.3. Salud en Chile	93
5.1.2.4. Sobre la estadística en Salud	94
5.1.3. Fuentes de datos para los ministerios de la presidencia	94
5.1.3.1. Ministerio de la Presidencia en Chile	94
5.1.3.2. Ministerio de la Presidencia en Costa Rica	95
5.1.3.3. Ministerio de la Presidencia en Uruguay	95
5.1.4. Algunas aclaraciones sobre la toma de datos	96
5.1.4.1. Dificultad de comparación	96
5.1.4.2. Regla general de selección	97
5.2. Resultados y análisis	98
5.2.1. Análisis global	147
5.2.2. Tablas de observaciones	148
6. Conclusiones	149
Bibliografía	150
Anexos	160
A. Cuestionario aplicado a los sitios	161
B. Imágenes de portales de datos abiertos	168
C. Literatura de dimensiones de datos	171

Índice de Tablas

2.1. Formatos de publicación de datos más utilizados y sus características.	14
2.2. Resumen de los hitos hacia la transparencia y los datos abiertos en Chile, Costa Rica y Uruguay	29
3.1. Definiciones de dimensiones de datos abierto	55
4.1. Ejemplo de conjunto de datos.	70
5.1. Cantidad de conjuntos de datos evaluados en Chile	86
5.2. Cantidad de conjuntos de datos evaluados en Costa Rica	87
5.3. Cantidad de conjuntos de datos evaluados en Uruguay	88
5.4. Cantidad de conjuntos de datos evaluados por gráfico	89

Índice de Ilustraciones

1.1. Cronograma propuesto para el desarrollo de la tesis, de junio del 2017 a abril del 2018	4
2.1. Relación entre datos, información, conocimiento y sabiduría, según el esquema de Ackoff.	7
2.2. Propuesta de clasificación de los principios de datos abiertos, según los ejes técnico, legal y político	13
2.3. Esquema de desarrollo de cinco estrellas para datos abiertos.	16
2.4. Flujo hacia la transparencia	19
2.5. Interacción entre los factores que inciden en la publicación de datos abiertos	20
3.1. Escenario de captura de datos a través de un sensor.	34
3.2. Escenario de creación de datos a partir de otros datos.	35
3.3. Escenario de productor de información.	36
3.4. Resumen de las principales dimensiones de los usos y usuarios de datos abiertos.	37
3.5. Dimensiones propuestas por Wang y Strong, 1996	38
3.6. Dimensiones propuestas por Wang y Guarascio, 1991.	39
3.7. Dimensiones propuestas por Thomas C. Redman.	40
3.8. Categorización de propiedades de datos en la literatura, parte 1.	41
3.9. Categorización de propiedades de datos en la literatura, parte 2.	42
3.10. Matriz de expectativas de usuarios propuestas y las funcionalidades de los portales Web	43
3.12. Estructura de las dimensiones de calidad de datos en portales Web planteada por Caro y otros	44
3.13. Clasificación de características de calidad de datos propuesta por la el estándar ISO/IEC 25012	45
3.14. Estructura de las dimensiones de calidad de datos en portales Web planteada en el SPDQM por Moraga y colegas	46
3.15. Dimensiones y subdimensiones (niveles) propuestas por Abella	47
3.16. Problemas en los datos abiertos detectados por Vetrò y colegas	48
3.17. Métricas para datos abiertos propuestas por Vetrò y colegas	49
3.18. Línea de tiempo de literatura de calidad de datos	51
3.19. Distribución de las dimensiones de acuerdo a sus usos y usuarios.	58
3.20. Dimensiones compartidas y específicas según a sus usos y usuarios	59
4.1. Indicadores/preguntas creados para la dimensión ciudadano y la sub dimensión valoración, en ITAI	64

4.2.	Nombres y descripciones de los metadatos obligatorios, propuestos en la Norma técnica para publicación de datos abiertos en Chile	65
4.3.	Sub variables e indicadores para la variable “ejecución del presupuesto”	66
5.1.	Complejidad de homologación en los Ministerios de Chile, Costa Rica y Uruguay. Siglas: Instituto Costarricense de Turismo (ICT), Instituto Costarricense del Deporte y la Recreación (ICODER)	85
5.2.	Clasificación del apartado de Matrícula y Cobertura de la ANEP. La primera columna muestra el nombre de las secciones y sus archivos y en las siguientes columnas los nombres de los apartados donde están presente la sección y los archivos, parte 2	90
5.3.	Clasificación del apartado de Matrícula y Cobertura de la ANEP. La primera columna muestra el nombre de las secciones y sus archivos y en las siguientes columnas los nombres de los apartados donde están presente la sección y los archivos, parte 2	91
5.4.	Publicaciones del 2016 y 2017, sección de estadística del ministerio de educación de Costa Rica	98
5.5.	Sección de estadística del ministerio de educación de Chile	99
5.6.	Gráfico del Ministerio de Educación de Chile, sección de transparencia	100
5.7.	Gráfico del Ministerio de Educación de Costa Rica, sección de transparencia	102
5.8.	Gráfico del Ministerio de Educación de Uruguay, sección de transparencia	104
5.9.	Gráfico del Ministerio de Educación de Chile, sección de estadística	106
5.10.	Gráfico de la Administración Nacional de Educación Pública de Uruguay	108
5.11.	Gráfico del Ministerio de Educación de Uruguay, sección de estadística	110
5.12.	Gráfico del Ministerio de Educación de Uruguay, sección de estadística en PDF	112
5.13.	Gráfico del Ministerio de Educación de Costa Rica, sección de estadística en hoja de cálculo	114
5.14.	Gráfico del Ministerio de Educación de Costa Rica, sección de estadística	116
5.15.	Gráfico del Ministerio de Salud de Chile, sección de transparencia.	118
5.16.	Gráfico del Ministerio de Salud de Costa Rica, sección de transparencia.	120
5.17.	Gráfico del Ministerio de Salud de Uruguay, sección de transparencia.	122
5.18.	Gráfico del Ministerio de Salud de Chile, sección de estadística. Evaluación de datos sobre el síndrome de inmunodeficiencia adquirida (SIDA)	124
5.19.	Gráfico del Ministerio de Salud de Costa Rica, sección de estadística. Evaluación de datos sobre el síndrome de inmunodeficiencia adquirida (SIDA)	126
5.20.	Gráfico del Ministerio de Salud de Uruguay, sección de estadística. Evaluación de datos sobre el síndrome de inmunodeficiencia adquirida (SIDA)	128
5.21.	Gráfico del Ministerio de Salud de Chile, sección de estadística. Evaluación de datos correspondientes a los Indicadores básicos de salud (IBS)	130
5.22.	Gráfico del Ministerio de Salud de Costa Rica, sección de estadística. Evaluación de datos correspondientes a los Indicadores básicos de salud (IBS)	132
5.23.	Gráfico del Ministerio de Salud de Uruguay, sección de estadística. Evaluación de datos correspondientes a los Indicadores básicos de salud (IBS)	134
5.24.	Gráfico del Ministerio de Salud de Chile, sección de estadística. Evaluación de datos correspondientes a Enfermedades de notificación obligatoria (ENO)	136

5.25. Gráfico del Ministerio de Salud de Costa Rica, sección de estadística. Evaluación de datos correspondientes a Enfermedades de notificación obligatoria (ENO)	138
5.26. Gráfico del Ministerio de Salud de Uruguay, sección de estadística. Evaluación de datos correspondientes a Enfermedades de notificación obligatoria (ENO)	140
5.27. Gráfico del Ministerio de Presidencia de Chile, sección de transparencia. . . .	142
5.28. Gráfico del Ministerio de Presidencia de Costa Rica, sección de transparencia.	144
5.29. Gráfico del Ministerio de Presidencia de Uruguay, sección de transparencia. .	146

Capítulo 1

Introducción

Es común escuchar a personas hablar de datos e información como una misma cosa. Sin embargo, esto no es correcto ya que un dato es un símbolo o un conjunto de símbolos que representan una característica de un objeto o de un hecho. Mientras, la información es un conjunto de datos que se han ordenado y complementado de manera que construyen un mensaje y que cambia el estado de conocimiento de quien lo recibe. De esta manera podemos decir que la información a diferencia de los datos puede ser interpretada, pues tiene semántica. Aunque se suelen confundir mucho estos términos es necesario tener claro que el dato por sí solo no comunica nada, mientras que la información cambia el estado de conocimiento de quien la recibe.

Los datos que se encuentran disponibles abiertamente a las personas y con la libertad para ser utilizados, reutilizados y redistribuidos son considerados como “datos abiertos”. Esta denominación es parte de una filosofía de apertura, que motivada por las libertades del software libre se ha venido extendiendo en los últimos años. Ha alcanzado el ámbito musical, los textos científicos e incluso actualmente a los datos e información del sector público, a través de un movimiento denominado Gobierno Abierto.

El Gobierno Abierto y la transparencia en los gobiernos están íntimamente ligadas. La transparencia se puede entender como aquella característica del gobierno que permite a los ciudadanos y a quien lo desee ver con claridad sus acciones, sus planes y el manejo de los recursos. Por lo que las autoridades deberán promover la publicación de datos e información a través de medios físicos y digitales. Por otro lado, el Gobierno Abierto surge con el fin de promover el desarrollo y la innovación en los países. En el Gobierno Abierto la publicación de datos trasciende la idea de sólo transparentar la gestión y considera la oportunidad de brindar a los ciudadanos datos con las características técnicas y jurídicas que les permita reutilizarlos para mejorar la toma de decisiones, innovar y generar bienestar.

Esta política de datos abiertos es una tendencia mundial y los gobiernos de América Latina se han integrado a esta iniciativa de Gobierno Abierto y han empezado a liberar datos tanto para buscar la transparencia como para promover el desarrollo de sus países. Es en este escenario donde nace el presente trabajo, con la idea de estudiar las características que debieran tener estos datos y ayudar a mejorar su publicación, a fin de colaborar con los

gobiernos para que la publicación de datos sea realmente de utilidad a sus usuarios. Para esto se estudió y comparó los casos de la publicación de datos de los poderes ejecutivos (llamados informalmente gobiernos) de Chile, Costa Rica y Uruguay.

El presente trabajo se focaliza en el estudio de la publicación de datos de gobierno y el desarrollo de una metodología que permita estimar el nivel de apertura, capacidad de reutilización y acceso de los datos que publican esos gobiernos. Para ello se estudiaron e identificaron los tipos de usuarios de datos abiertos y los posibles usos que estos usuarios les pueden dar. Con esto se estudió la literatura de dimensiones de datos y se propuso una metodología genérica y métricas que permitan evaluar el nivel de apertura, capacidad de reutilización y acceso de los conjuntos de datos que publican los gobiernos. Además de un esquema que ayuda a conocer a qué tipo de uso y usuario está dirigido un conjunto de datos.

Es importante aclarar, dado que se puede interpretar de diversas maneras, qué entendemos por apertura, capacidad de reutilización y accesibilidad. Con *apertura* nos referimos a las características legales y técnicas que los conjuntos de datos deben tener. O sea, que tenga licencia abierta, formato abierto, entre otras cosas. La *capacidad de reutilización* se mide considerando los usos que se le pueden dar a los conjuntos de datos. Es decir, si los datos pueden ser solamente consumidos, o si por el contrario, tienen características que permitan a los usuarios producir datos a partir del conjunto evaluado. También se refiere a si son reutilizables por un usuario técnico o no técnico. El *acceso* fue evaluado considerando el tamaño de los archivos donde se encuentra el conjunto de datos y con ello si es fácil de manipular. También considera si existe un catálogo de datos que permita acceder con facilidad a los conjuntos.

Se eligió como casos de estudio los gobiernos de: Chile, Costa Rica y Uruguay por las siguientes razones. Son tres países que tienen un muy buen desarrollo en la publicación de datos de gobierno. Son tres países que tienen tradiciones políticas similares y son reconocidas como democracias sólidas de la región. Por otro lado, son los países menos corruptos de América Latina según el Índice de Percepción de la Corrupción (IPC) publicado por la organización Transparencia Internacional en 2015 y el 2017. Además poseen una división político administrativa similar, es decir, se encuentran divididos en provincias o departamentos y estos a su vez en comunas (también llamados cantones o municipios).

Esta tesis se propuso los siguientes objetivos y metodología:

1.1. Objetivo general

- Diseñar e implementar una metodología genérica y métricas que permitan evaluar el nivel de apertura, capacidad de reutilización y accesibilidad de los datos de gobierno, particularmente los de transparencia, considerando como casos de estudio Chile, Costa Rica y Uruguay.

1.2. Objetivos específicos

- Estudiar las metodologías de evaluación de información de portales gubernamentales desde una perspectiva de datos
- Revisar la literatura sobre datos abiertos de gobierno particularmente transparencia
- Diseñar una metodología y métricas para evaluar los datos abiertos de gobierno particularmente transparencia
- Evaluar la apertura, capacidad de reutilización y accesibilidad de los datos de transparencia de los gobiernos de: Chile, Costa Rica y Uruguay usando la metodología diseñada. Conceptualizar y comparar los sitios de los gobiernos mencionados

1.3. Plan y metodología

Los pasos a seguir para realizar este trabajo se plantearon de la siguiente manera:

1. **Revisión bibliográfica sobre transparencia y datos abiertos:** Este trabajo consiste en estudiar y analizar las publicaciones sobre índices y leyes de transparencia, teoría de datos abiertos y formas de publicación de éstos, propiedades de los datos, entre otras.
2. **Revisión de sitios de gobierno:** En esta revisión se van a obtener y estudiar los datos pertenecientes a los ministerios de los gobiernos de Chile, Costa Rica y Uruguay. A los cuales posteriormente se le aplicará la metodología de evaluación.
3. **Revisión bibliográfica sobre métodos y métricas:** En esta etapa se realizará la síntesis y análisis de las metodologías y métricas de evaluaciones de portales web de gobiernos.
4. **Diseño y aplicación de parámetros y métricas que permitan realizar la revisión/evaluación de sitios:** Esto considerando las propiedades de los datos y la conceptualización y el ordenamiento de los temas en los sitios.
5. **Escribir informe:** Después de evaluar los resultados de la aplicación de la metodología en los sitios gubernamentales se finalizará la documentación.

1.4. Estructura del presente documento

Este texto documenta el desarrollo de la tesis y se encuentra compuesto por cinco capítulos, los cuales se describen a continuación.

1. **Introducción:** El capítulo que usted se encuentra leyendo contiene la descripción del trabajo realizado. En particular, los objetivos, la metodología, el cronograma y la descripción de la estructura del documento.

Actividades metodológicas	julio	agosto	septiembre	octubre	noviembre	diciembre	enero	febrero	marzo	abril
1	■	■	■							
2				■	■					
3						■	■			
4							■	■	■	
5								■	■	■

Figura 1.1: Cronograma propuesto para el desarrollo de la tesis, de junio del 2017 a abril del 2018

Fuente: Creación propia

2. **Datos abiertos, Gobierno Abierto y Transparencia:** Acá se habla sobre datos, información y conocimiento, qué es y cuál es la diferencia entre ellos. Introduce el tema de ecosistemas de datos y datamediarios. Expone las características de datos abiertos y el surgimiento de Gobierno Abierto. Explica qué es transparencia y cómo se originó. Se analizan los principales hitos de transparencia en Chile, Costa Rica y Uruguay.
3. **Producción, usos, usuarios y dimensiones de los datos abiertos:** Este capítulo contiene la definición y clasificación de los usuarios de datos abiertos considerando la teoría estudiada. Se analizan los usos que le dan estos usuarios a los datos y las dimensiones que le interesan a cada usuario.
4. **Métricas e indicadores para datos abiertos:** En este apartado se definen métricas e indicadores para proponer una metodología que permita evaluar la apertura de los conjunto de datos en las secciones de transparencia de acuerdo a los usos y usuarios de los datos.
5. **Análisis y resultados:** Acá se presentan los resultados de las evaluaciones realizadas en los portales de transparencia y de datos abiertos de los gobiernos de Chile, Costa Rica y Uruguay.

Capítulo 2

Datos abiertos, Gobierno Abierto y Transparencia

En este capítulo se aborda el tema de datos abiertos aclarando la diferencia entre datos, información y conocimiento. Luego se exponen los principios aplicados a los datos publicados por los gobiernos abiertos, los definidos por la carta internacional de datos abiertos y el *Open Data White Paper*, publicado por Reino Unido. A continuación se describe el modelo de evaluación de cinco estrellas propuesto por Tim Berners-Lee y se explica brevemente los términos de licencia de los datos abiertos. Esta sección finaliza con un repaso de la literatura de usuarios de datos abiertos haciendo énfasis en los intermediarios de datos. Finalmente, se estudia la analogía de los ecosistemas de datos abiertos.

En la segunda parte del capítulo se estudia la noción de transparencia, qué la precede y qué acciones han motivado su creación. Luego se comentan los principales hitos de transparencia en Chile, Costa Rica y Uruguay. Se analiza tanto la legislación nacional como internacional, y la presión que ejercen las organizaciones internacionales sobre estos países para promover la creación de mecanismos que ayuden a mitigar la corrupción, aumentar la transparencia y promover el desarrollo. Al finalizar el capítulo el lector encontrará una muy breve revisión de la literatura relacionada con transparencia en la región latinoamericana.

2.1. Datos

En esta sección se estudian los conceptos de dato, información y conocimiento. Además, expone el rol de los datamediarios en el consumo y la producción de datos. Concluye haciendo una revisión sobre la literatura de ecosistemas de datos y como los usuarios y usos estudiados en este trabajo encajan en estos ecosistemas.

2.1.1. Datos, información y conocimiento

Para abordar el tema de los datos abiertos se comenzará aclarando la diferencia entre datos e información. Para Ackoff [3] los datos son “símbolos que representan las propiedades de los objetos y eventos”. Según Talledo [92] un dato “es una representación simbólica de carácter lingüístico o numérico de un atributo o variable cuantitativa o cualitativa”. Para Abellán [2] un dato es “una representación simbólica de un conocimiento o una situación, pero sin ningún sentido semántico, que describe situaciones y hechos sin transmitir ningún mensaje.” Aunque se suelen confundir mucho los términos datos e información, es necesario tener claro que el dato por sí solo no comunica nada, mientras que la información cambia el estado de conocimiento de quien la recibe.

La definición de Ackoff [3] para información dice así: “La información consiste en datos procesados, el procesamiento dirigido a aumentar su utilidad”. Con el fin de aclarar más la diferencia entre datos e información, consideraremos la definición de información propuesta por Velázquez [96] “La información es un conjunto organizado de datos procesados, que constituyen un mensaje que cambia el estado de conocimiento del sujeto o sistema que recibe dicho mensaje”. De esta manera podemos decir que la información a diferencia de los datos puede ser interpretada, pues tiene semántica.

Por otra parte, el tema del conocimiento ha sido abordado desde diferentes perspectivas. Fritz Machlup lo examina como recurso económico [57], [58], [59], hace un breve repaso desde el punto de vista filosófico, del proceso de aprendizaje educativo y un acercamiento a la sociedad del conocimiento [63]. Con una posición más cercana a la ciencias de la computación como lo es la Visualización, Chen y compañía [20] definen datos como “representaciones computarizadas de modelos y atributos de entidades reales o simuladas”, información como “datos que representan los resultados de un proceso computacional, como el análisis estadístico, para asignar significados a los datos, o las transcripciones de algunos significados asignados por los seres humanos” y conocimiento como “Datos que representan los resultados de un proceso cognitivo simulado por computadora, como la percepción, el aprendizaje, la asociación y el razonamiento, o las transcripciones de algunos conocimientos adquiridos por los seres humanos”. Russell Ackoff [3] presenta con un lente más cognitivo, y amplía el proceso a datos, información conocimiento y sabiduría. Para Ackoff la información “está contenida en descripciones, respuestas a preguntas que comienzan con palabras tales como quién, qué, cuándo, dónde y cuántas. El conocimiento se transmite mediante instrucciones, respuestas a preguntas prácticas”. Por su parte Teskey [93] propone que la distinción entre datos, información y conocimiento se debe basar en las siguiente nociones (cito textualmente):

- Los datos son el resultado de la observación directa de eventos, o sea, valores de atributos de objetos.
- La información son colecciones estructuradas de datos, conjuntos de datos, relaciones entre datos.
- El conocimiento son modelos del mundo que pueden ser creados o modificados por nueva información.

La figura 2.1 muestra una pirámide con la relación entre: datos, información, conocimiento y sabiduría. Los datos son la base de pirámide, es la entidad más elemental de la cadena;

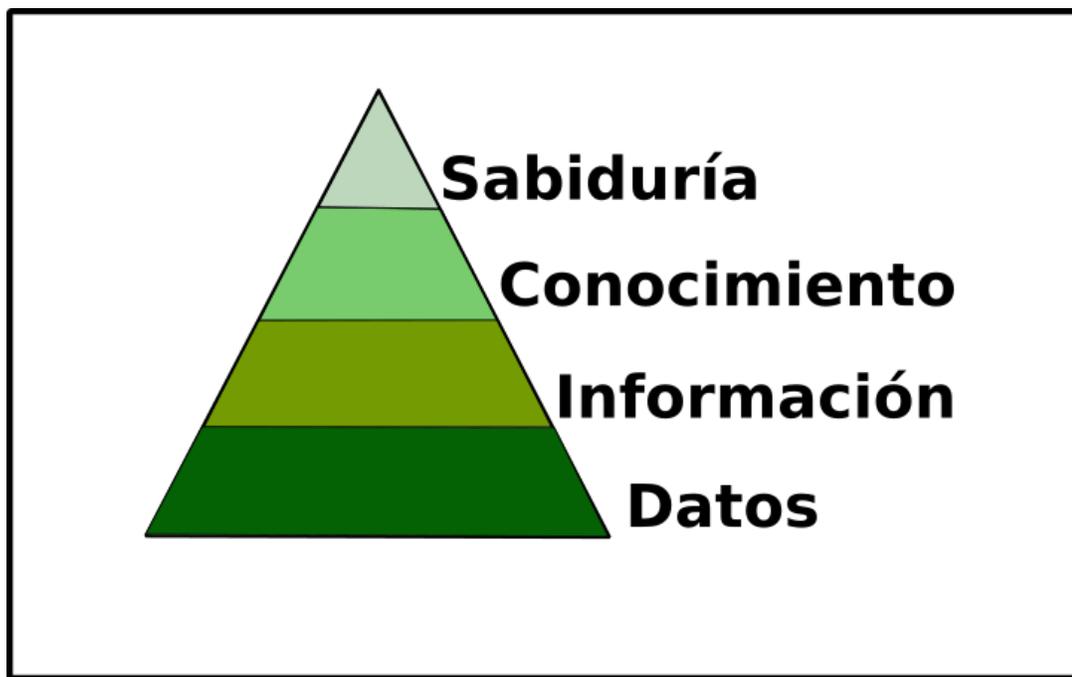


Figura 2.1: Relación entre datos, información, conocimiento y sabiduría, según el esquema de Ackoff.

Fuente: Imagen creada a partir de [3]

la información está compuesta por datos relacionados e interpretados; luego sigue el conocimiento, que más que un conjunto de información dispersa, involucra procesamiento humano, que puede estar plasmado incluso en rutinas de una organización. Finalmente, en la cúspide se ubica la sabiduría, como la más difícil de alcanzar, que sería “la habilidad de ver con anticipación las consecuencias de las acciones actuales”.

Watson, citado por Guerrero y Pino [43] sugiere las siguientes definiciones para datos, información y conocimiento. (Cito textualmente de Guerrero y Pino)

Los datos son hechos que no se han analizado ni resumido (datos brutos). La información se procesa en forma significativa, aunque “significativa” depende del contexto; por ejemplo, la información de una persona puede ser información de otra persona. Finalmente, el conocimiento se explica como la capacidad de usar información, es decir, saber qué información se requiere y cuál es el significado de dicha información.

2.1.2. Productores/Usuarios/Intermediarios de datos o Datamedia-rios

Un *datamediario* o *intermediario de datos* es un individuo u organización que obtiene datos del Sector Público o Privado, los analiza, procesa y los entrega con valor agregado, es decir, crea productos o servicios útiles a sus clientes. Esto incluye tanto a quienes se dedican por completo a esta finalidad, como aquellas personas que lo hacen de forma no constante y

a las entidades que poseen solo una parte de la organización dedicada a esto.

En la literatura se encuentra un primer acercamiento a los intermediarios electrónicos en un estudio exploratorio presentado en 1997 por Bailey y Bakos [7] en el cual se exploran trece casos de estudios y se evidencian los roles emergentes de los intermediarios en el sector de comercio electrónico. Por su parte, Hagel [44] destaca la labor del intermediario como aquel que captura datos de consumidores y desarrolla perfiles detallados de clientes individuales para uso de terceros proveedores seleccionados. En este sentido, en el capítulo cuatro del libro *Data Protection Law in Singapore: Privacy and Sovereignty in an Interconnected World* [21] Seng analiza el aspecto legal de la captura de datos realizada por los intermediarios. Además describe algunos de los negocios relacionados con los datos que en la actualidad realizan algunos datamediarios como Google, Facebook, Pinterest, entre otras.

Los datamediarios no siempre obtienen datos personales y sus actividades no se encuentran limitadas a los fines lucrativos. Treuhaft en su informe [94] analiza el trabajo de los datamediarios cívicos que utilizan las tecnologías para promover el desarrollo comunitario de las zonas urbanas facilitando la comunicación de datos principalmente a través de Internet.

Ante el diluvio de datos que se vive en esta época. Ho [47] estudia el caso de cinco industrias: energía y utilidad, farmacéutica, financiera, editorial y logística e identifica tres factores comunes que deben abordar los datamediarios para acercarse al buen manejo de los datos. Estos factores son: la cantidad, el formato y la calidad de los datos.

El papel que juegan los datamediarios con respecto a los datos es múltiple, pues dado que parten de los datos para poder entregar un producto, son consumidores de datos. Por otra parte, por su naturaleza son intermediarios y a la vez se pueden convertir en productores de datos, si lo que ofrecen después de agregar valor a los datos también es un producto de datos.

Janssen y Zuiderwijk [48] identifican dos variables que diferencian los modelos de negocio de los intermediarios: el nivel de acceso de datos y nivel de diálogo. El primero tiene una relación con la capacidad técnica del usuario al que va dirigida. Por ejemplo, si la vista a los datos está definida de forma que facilita su uso interpretativo, estará dirigido a un usuario no técnico; el caso en que los datos estén más granulares indica su uso a usuarios más técnicos. El segundo nivel se puede enmarcar en la categoría de usuarios no técnicos, que pretenden comprender los datos o productos con valor agregado.

2.1.3. Ecosistemas de datos

La idea de realizar una analogía entre los ecosistemas naturales y el contexto de los datos no es nueva. Algunos trabajos ya han realizado un acercamiento entre los ecosistemas y los actores en un contexto digital [75] y específicamente en el caso de datos [67],[79] [45], [83], y de datos en el entorno financiero [76]. Según Nardi [67] un ecosistema es un “sistema de personas, prácticas, valores y tecnologías en un entorno local particular”. Por ejemplo, la oficina de estadística del Ministerio de Salud puede establecer vínculos con organizaciones para brindar datos de salud de forma periódica, y estas organizaciones a su vez pueden procesar y añadir valor a los datos. Así muchas oficinas de diferentes ministerios pueden

interactuar con muchas organizaciones creando múltiples ecosistemas.

Para comprender mejor los ecosistemas se presenta a continuación un ejemplo real. Pinterest es una red social basada en imágenes, que permite a sus usuarios guardar imágenes, categorizarlas en diferentes tableros, buscar y seguir los tableros de los usuarios si tienen los mismos gustos, etc. El sitio se puede usar de forma “gratuita”, pero a cambio recopila datos de los gustos, disgustos, intereses, relaciones entre los comportamientos, entre otros. Estos valiosos datos permiten realizar estudios de mercado, crear perfiles de usuarios con datos importantes para empresas de marketing u otras redes sociales para complementar perfiles y otras acciones. Finalmente esto puede redundar en que la publicidad en los navegadores sea más asertiva. Este conjunto de actores que incluyen los usuarios, las empresas que procesan los datos, los que los compran y agregan más valor y revenden, quienes crean y colocan los anuncios a los usuarios de acuerdo a los datos capturados, todos estos se encuentran en el sistema, o más claramente, en un ecosistema de datos.

Esta analogía con los ecosistemas nos permite considerar a las organizaciones que requieren obtener su materia prima (datos) de lo que otros producen (es decir, de una base de datos ya creada) y relacionarlos con los seres consumidores. En la figura 3.1 del capítulo 3 el escenario nos muestra un productor de datos que no requiere de nada más que el ambiente para producir sus datos. Más que un tema de ambiente se puede considerar que no requiere de lo que otros produzcan, y quizá esto permita relacionarlo con los productores en un ecosistema natural. Lo que sí es bien claro es que tenemos *consumidores* y *productores*, y que de los consumidores los podemos encontrar de primer nivel, segundo nivel y hasta nivel n , dado que los consumidores pueden seguir usando y procesando lo que otros han procesado.

2.2. Datos abiertos

En este apartado se introduce al lector al tema de datos abiertos, además se describen: sus orígenes, sus principios, sus características y su relación con el movimiento open. También muestra la propuesta de evaluación de datos abiertos de Berners-Lee. Se expone la definición y la importancia de los datos enlazados. Esta sección finaliza con el estudio del concepto de Gobierno Abierto.

2.2.1. La tendencia “Open” y los datos abiertos

Para dar un mejor sentido al término “abierto” se considera la definición ofrecida por la Fundación Conocimiento Abierto [72] (en inglés, *Open Knowledge Foundation*, OKF) aplicada al conocimiento: “es abierto si cualquiera es libre para acceder a él, usarlo, modificarlo y compartirlo bajo condiciones que, como mucho, preserven su autoría y su apertura”. Esta definición se corresponde con la de “open” cuando se refiere al software como en la definición de *Open Source*, y con la de “free” aplicadas a las obras culturales libres. La página de obras culturales libres [41] expone la aplicación de las obras abiertas en contextos como: la música, libros, y otras obras culturales; considerando las libertades y referencias a las licencias que

fomentan estas libertades.

Licencia abierta, acceso y formato abierto son parte de los requisitos que debe tener una obra para considerarse abierta. Cuando se menciona licencia se refiere a las condiciones legales bajo las cuales está disponible la obra. Según la Fundación Conocimiento Abierto [72] para que una licencia sea abierta debe permitir irrevocablemente lo siguiente:

- **Uso:** Debe autorizar el libre uso de la obra.
- **Redistribución:** Debe permitir la distribución de la obra en sí misma o como parte de una colección compuesta por obras de diversas fuentes.
- **Modificación:** Debe permitir la creación de obras derivadas de la obra licenciada y la distribución de tales derivados en los mismos términos que la obra original licenciada.
- **Separación:** Si una obra es originalmente distribuida como una colección de obras, se debe permitir que cualquier parte de la obra sea libremente usada, distribuida o modificada por separado de la colección original. Todas las partes que reciben una distribución de cualquier parte de una obra en los términos de la licencia original deben tener los derechos que están garantizados en relación a la obra original.
- **Compilación:** Se debe permitir distribuir la obra junto con otras obras distintas sin poner restricciones sobre las otras obras.
- **No discriminación:** La licencia no debe discriminar a ninguna persona o grupo.
- **Propagación:** Los derechos ligados a la obra deben aplicarse a toda persona a quien le sea distribuida sin necesidad de aceptar ningún término legal adicional.
- **Aplicación para cualquier propósito:** La licencia no debe restringir a nadie a hacer uso de la obra en un campo específico de aplicación o trabajo.
- **Sin coste:** La licencia no debe imponer en sus condiciones ningún esquema de tarifa, compensación o remuneración económica.

La licencia también puede contener condiciones de limitación como las que se citan a continuación: Reconocimiento, integridad, compartir igual, entre otras.

Ya se ha dejado clara la definición de datos y se ha realizado un acercamiento al término “abierto”, de manera que ya se abre paso a la interpretación de la fase “datos abiertos”. En el *Manual de datos abiertos* [71], se encuentra su definición de la siguiente manera: “Los datos abiertos son datos que pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier persona, y que se encuentran sujetos, cuando más, al requerimiento de atribución y de compartirse de la misma manera en que aparecen”.

La página de Datos Abiertos de Gobierno [74] propone 8 principios aplicados a los datos publicados por los gobiernos como parte de las acciones en búsqueda de la transparencia y del Gobierno Abierto. A continuación se describen cada unos de ellos (cito parafraseado):

1. **Completos:** Todos los datos se deben poner a disposición del público. Los datos no deben estar sujetos a limitaciones vigentes de privacidad, seguridad o privilegios.
2. **Primarios:** Los datos se debe hacer disponibles de la forma más granular posible, no en forma agregada o modificada.

3. **Oportunos:** Los datos deben estar disponibles lo más pronto posible, a fin de preservar su valor.
4. **Accesibles:** Los datos deben estar disponibles para la mayor cantidad de usuarios y para la mayor cantidad de propósitos posibles.
5. **Procesables por máquina:** Los datos deben estar razonablemente estructurados para permitir que estos sean procesados automáticamente. Por ejemplo, separados por tabulador. (Ver tabla 2.1)
6. **No discriminatorios:** Los datos deben estar disponibles para cualquier persona, sin necesidad de registro.
7. **Formato abierto:** Los datos deben estar en un formato sobre el cual ninguna entidad tenga el control exclusivo. (Ver tabla 2.1)
8. **Licencia abierta:** Los datos no deben estar sujetos a ningún derecho de autor, patentes, marcas o reglamento de secreto comercial.

Estos ocho principios fueron parte del producto del taller de Gobierno Abierto realizado los días 7 y 8 de diciembre del 2007 en Sebastopol, California, evento que congregó a 30 representantes de organizaciones interesados/defensoras del Gobierno Abierto [74].

Por otro lado, la *Carta Internacional de Datos Abiertos* [73] define datos abiertos como: “datos digitales que son puestos a disposición con las características técnicas y jurídicas necesarias para que puedan ser usados, reutilizados y redistribuidos libremente por cualquier persona, en cualquier momento y en cualquier lugar”. Los adherentes a esta carta acuerdan seguir seis principios que sientan las bases para el acceso a los datos y para su publicación y uso. Dichos principios indican que los datos deben ser (cito textualmente):

1. Abiertos por defecto
2. Oportunos y exhaustivos
3. Accesibles y utilizables
4. Comparables e interoperables
5. Para mejorar la gobernanza y la participación ciudadana
6. Para el desarrollo incluyente y la innovación

En junio del 2012, Francis Maude, Ministro de la Oficina del Gabinete y Paymaster General del Reino Unido publicó un documento llamado Open Data White Paper [61], en el cual se describen 14 principios elaborados por la Junta de Transparencia del Sector Público y revisados como resultado de consultas, a continuación se citan textualmente:

1. La política y la práctica de datos públicos estarán claramente impulsadas por el público y las empresas que desean y usan los datos, incluido qué datos se publican cuándo y en

qué forma

2. Los datos públicos se publicarán en forma reutilizable y legible por máquina
3. Los datos públicos se publicarán bajo la misma licencia abierta que permite la reutilización gratuita, incluida la reutilización comercial
4. Los datos públicos estarán disponibles y serán fáciles de encontrar a través de un único punto de acceso en línea fácil de usar (www.data.gov.uk)
5. Los datos públicos se publicarán utilizando estándares abiertos y siguiendo las recomendaciones pertinentes del World Wide Web Consortium (W3C)
6. Se publicarán datos públicos de diferentes departamentos sobre el mismo tema en los mismos formatos estándar y con las mismas definiciones
7. Los datos públicos en los que se basan los sitios web del Gobierno se publicarán en forma reutilizable
8. Los datos públicos serán oportunos y de grano fino
9. Libere datos rápidamente y luego trabaje para asegurarse de que estén disponibles en formatos estándar abiertos, incluidos los enlaces formularios de datos
10. Los datos públicos estarán disponibles gratuitamente para su uso de cualquier manera legal
11. Los datos públicos estarán disponibles sin solicitud o registro, y sin requerir detalles del usuario
12. Los organismos públicos deberían alentar activamente la reutilización de sus datos públicos
13. Los organismos públicos deberían mantener y publicar inventarios de sus tenencias de datos
14. Los organismos públicos deberían publicar metadatos relevantes sobre sus conjuntos de datos, y deberían estar disponibles a través de un único punto de acceso en línea; y deben publicar descripciones de apoyo del formato, procedencia y significado de los datos

La Organización para la Cooperación y el Desarrollo Económicos publicó en el 2007 un documento llamado “Principios y directrices de la OCDE para el acceso a los datos de investigación de la financiación pública” [39] donde describen principios que deben poseer los datos de investigación que se producen con dineros del gobierno. Entre sus directrices indican que deben ser abiertos, señalando una semejanza con los datos abiertos de gobierno, que en el fondo lo son.

En la figura 2.2 se muestra una propuesta de clasificación de los principios creados por

	Técnico	Legal	Político
8 principios 2007, Opengovdata	2-Primario	8-Licencia abierta	1-Completo
	4-Accesible		3-Oportuno
	5-Procesable por máquina		
	6-No discriminatorio		
	7-No propietario		
OCDE, Datos abiertos de investigación	B-Flexible		A-Apertura
	C-Transparencia	D-Conformidad Legal	F-Responsabilidad formal
	H- Interoperabilidad	E- Propiedad Intelec.	K. Eficiencia
	I. Calidad		L. Responsabilidad
	J. seguridad		M. Sostenibilidad
National Science Foundation (NSF)	2-Accesible	4-Reutilizable	1-Público
	3-Descrito		6-Oportuna
	5-Completo		7-Contacto
Carta Internacional de Datos Abiertos	1. Abiertos por defectos.		2. Oportunos y exhaustivos
	3. Accesibles y utilizables		5. Para mejorar la gobernanza y la participación ciudadana
	4. Comparables e interoperables		6. Para el desarrollo incluyente y la innovación

Figura 2.2: Propuesta de clasificación de los principios de datos abiertos, según los ejes técnico, legal y político
Fuente: Creación propia

las iniciativas de datos abiertos, entre ellas la Fundación nacional para la ciencia (NSF por sus siglas en inglés) [68], la OCDE [39], Opengovdata [74], y la carta internacional de datos abiertos [73]. Los principios se han clasificado en tres dimensiones: técnico, legal y político. La primera hace referencia al área técnica de la computación, la segunda envuelve el tema legislativo tanto nacional como internacional y el último abarca el ámbito administrativo/-político.

En el conjunto de principios definidos por la página de datos abiertos de gobierno [74] se indica que los datos deben ser procesables por máquinas, de manera que al descargarse se puedan volver a utilizar. Algunos ejemplos de reutilización pueden ser: la creación de mapas y la obtención de estadísticas por medio de lenguajes como R u otros. También se muestra la importancia de que los datos estén en un formato sobre el cual ninguna entidad tenga el control exclusivo. En la tabla 2.1 se muestra un cuadro resumen con los formatos más utilizados y sus características. Respecto al formato PDF que se indica en esta figura es necesario mencionar que si se utilizan fuentes privativas (no libres, sujetas a derechos de autor) para crear el documento, entonces el documento deja de ser libre. Por ejemplo, si se usa un editor de texto como Microsoft Word y al crear el documento se utiliza una fuente con derechos de autor el documento PDF creado deja de ser libre, pues las fuentes incrustadas tienen derechos de autor.

Formato	¿Procesable por máquina?	¿Formato abierto?
Texto plano (.txt)	Sí	Sí
Valores separados por coma (.csv/.txt)	Sí	Sí
Lenguaje de Marcas de Hipertexto (.html/.htm)	Sí	Sí
Lenguaje de Marcado Extensible (.xml)	Sí	Sí
Marco de Descripción de Recursos (.rdf)	Sí	Sí
Hoja de cálculo LibreOffice (.ods)	Sí	Sí
Microsof Excel (.xls)	Sí	No
Microsof Word (.doc)	No	No
Formato de texto enriquecido (.rtf)	No	No
Formato de documento portátil(.pdf)	No	Sí
Archivo de imagen (.jpg, .tif)	No	Sí

Tabla 2.1: Formatos de publicación de datos más utilizados y sus características.
Fuente: Adaptación de la tabla presentada por [22]

2.2.2. Datos enlazados

La Web fue creada con base en tres elementos principales: el protocolo HTTP (Hypertext Transfer Protocol) mediante el cual se accede a los recursos, el lenguaje HTML (Hiper Text Markup Language) con el que se crea y relacionan los documentos y los identificadores de recursos URI (uniform resource identifier) que identifican los documentos de forma global.

Inicialmente la Web trabajaba solo con documentos que se relacionaban con otros documentos a través de hipervínculos, fue pensada y diseñada para que fuera leída solo por seres humanos. Sin embargo, con el pasar del tiempo creció enormemente y la necesidad de buscar en esta inmensa red y compartir datos ha llevado a los expertos a buscar la forma de trasformarla de tal manera que las máquinas mismas la pueda interpretar. Pero, ¿Qué quiere decir esto de transformar la Web para que sea interpretada por las máquinas? La idea es que a través de un lenguaje como RDF (Resource Description Framework) se puedan describir relaciones entre recursos y visto de alguna manera entre objetos del mundo real, de manera que tanto los seres humanos como las computadoras puedan comprender la relación entre los recursos.

Entonces, ¿Que son los Datos Enlazados? Bizer y compañía [12] define datos enlazados como “un conjunto de mejores prácticas para publicar y conectar datos estructurados en la

Web”. Ahora bien, los datos abiertos enlazados es la forma en que Berners-Lee [10] propone que se publiquen los datos abiertos. Esto facilita la relación entre los conjuntos de datos y los recursos, además permite que las computadoras puedan interpretar las relaciones. Por ejemplo, un conjunto de datos con los salarios de los empleados del Ministerio de educación de Chile se puede relacionar con la entidad Ministerio de educación de Chile y conocer datos sobre el ministerio o conocer datos sobre el país Chile, considerando que exista una relación descrita en RDF que une estas entidades.

Con el fin de motivar a los gobiernos a publicar datos abiertos y enlazados Timothy "Tim" John Berners-Lee [10] diseñó un esquema de calificación basado en estrellas. Esta propuesta es una escala creciente (de una a cinco estrellas) relativa a la forma en que se publican los datos. La figura 2.3 resume gráficamente el esquema que a continuación se describe.

- **Una estrella:** Si ofrece los datos en la web en cualquier formato, pero, con licencia abierta.
- **Dos estrellas:** Si dispone los datos estructurados procesables por máquina. Por ejemplo en Excel.
- **Tres estrellas:** Si entrega los datos estructurados procesables por máquina y en un formato libre. Observe la tabla 2.1 para conocer la propiedad de formato libre.
- **Cuatro estrellas:** Todo lo anterior, y además usa los estándares abiertos de la W3C (RDF y SPARQL) para identificar las cosas (URIs), así la gente puede apuntar a ellas.
- **Cinco estrellas:** Todo lo anterior, y además enlaza los datos a otros datos para proveer contexto.

Para enlazar los datos Berners-Lee [10] definió cuatro reglas.

1. Usar URIs para identificar cosas, para el caso en estudio sería para identificar los conjuntos de datos
2. Hacer uso de URIs basadas en el protocolo HTTP.
3. Utilizar URI desreferenciables, que al escribir una URI en el navegador o en un crawler, se obtenga información acerca de esa URI en formatos estándares (HTML/humanos, RDF/máquinas)
4. Enlazar los recursos

Como ya se mencionó anteriormente, los datos que se encuentran disponibles abiertamente a las personas y con libertad para ser utilizados, reutilizados y redistribuidos son considerados como “datos abiertos”. Esta denominación es parte de una filosofía de apertura, que motivada por las libertades del software libre se ha venido extendiendo en los últimos años. Ha alcanzado el ámbito musical, los textos científicos e incluso actualmente a los datos e información del sector público, a través de un movimiento denominado Gobierno Abierto.

En países como Reino Unido [31], Canada [19], Estados Unidos [54], Países bajos (Holanda) [49], la publicación de datos abiertos en forma granular tuvo un gran impulso en los años recientes. Esta forma de publicación tiene ventajas, pero al mismo tiempo limita la posibilidad

de analizar los datos de forma manual, pues la gran cantidad de datos que contiene un conjunto de datos hace imposible que los ciudadanos puedan interpretarlos con rapidez, ¿Qué puede hacer un ciudadano interesado en conocer sobre su gobierno con 20 megabytes de datos en formato CVS? Entonces, ¿para quién son esos datos? Esto nos hace pensar que existen diferentes formas de publicar los datos, para diferentes tipos de usuarios interesados en los datos.

Como ya usted debe suponer, existen interesados en los datos con capacidades técnicas y con los instrumentos necesarios para procesar estos datos y crear productos o servicios para los ciudadanos. Estos actores capaces de obtener datos como materia prima y crear nuevos datos, visualizaciones o productos de información son conocidos como *intermediarios de datos*, (en la sección de datos en la página 7 ya se abordó este tema). Los intermediarios de datos son algunos de los usuarios de datos abiertos. En la literatura reciente varios estudios se han ocupado del tema [89], [62], [60], [35], [87]. La importancia de los intermediarios para este trabajo radica en la necesidad de conocer quién usa los datos abiertos y conocer qué usos le dan. En la página 31 se aborda con mayor profundidad lo que respecta a los usuarios y usos de los datos abiertos.

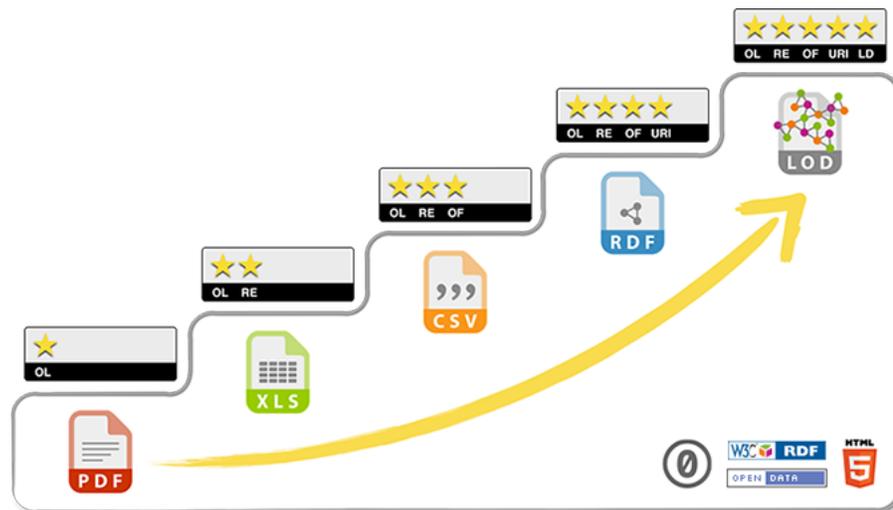


Figura 2.3: Esquema de desarrollo de cinco estrellas para datos abiertos.
Fuente: Tomado de: [46]

2.2.3. Gobierno Abierto

Se puede extraer del memorando [95] emitido por el entonces presidente de los Estados Unidos Barack Obama en 2009 que un Gobierno Abierto es un gobierno capaz de trabajar en conjunto para garantizar la confianza pública y establecer un sistema de transparencia, participación pública y colaboración.

Según Sandoval [86] el Gobierno Abierto debe entenderse como “una plataforma tecnológica institucional que convierta los datos gubernamentales en datos abiertos para permitir su uso, protección y colaboración por parte de los ciudadanos en los procesos de decisión pública, rendición de cuentas y mejoramiento de los servicios públicos”. La definición anterior fue

concebida en medio de conceptos que han surgido recientemente como gobierno digital, gobierno transparente y datos abiertos. También se puede entender como un nuevo paradigma o modelo de relación entre el gobierno y los ciudadanos caracterizado por multidireccionalidad, transparencia, y que posibilita crear y articular valor público desde y más allá de la burocracia gubernamental (Ramírez-Alujas citado por Cruz [29]).

Sandoval [86] brinda una aclaración teórica sobre los conceptos de Gobierno Abierto y transparencia, y en conjunto con Cruz [29] permite reconocer con más claridad qué es y qué no es Gobierno Abierto, considerando el surgimiento de conceptos similares como gobierno electrónico y gobernanza electrónica. Chiaretti [22] define Gobierno Abierto, datos abiertos y datos abiertos enlazados. También realiza un recorrido por los sitios Web de de transparencia gobierno Chile para dar a conocer el estado de avance de la publicación de datos abiertos. Además presenta una comparación de los avances en la entrega de datos abiertos entre Chile y otros países de la región.

Considerando la definición anterior y la analogía de Gobierno Abierto con un ecosistema se puede plantear que el Gobierno Abierto es un ecosistema socio-tecnológico en el que se relacionan productores e intermediadores de datos, usuarios técnicos y no técnicos, organizaciones públicas y privadas.

Por otro lado, Cruz [29] concibe el Gobierno Abierto como una filosofía política administrativa, un nuevo paradigma o modelo de interacción sociopolítica basada en los principios y valores de transparencia, rendición de cuentas y del uso de la tecnología; que posiciona a los ciudadanos como centro de atención y de prioridad.

Es importante aclarar que los datos abiertos no son lo mismo que Gobierno Abierto, que los gobiernos abiertos publican datos abiertos, pero que no es lo único que hacen y el término no se limita a eso. Por otro lado, los datos abiertos son aquellos que cumplen con las condiciones antes citadas, y que hasta organizaciones no gubernamentales, e incluso privadas pueden publicar datos abiertos. Por ejemplo, Wikidata, Banco mundial, etc.

Finalmente, se han planteado diferentes ideas para mejorar la exposición de datos abiertos de gobierno. La Red cívica de analistas [23] ofrece ocho consejos para mejorar las capacidades de los portales de gobierno transparente, que a continuación se resumen:

1. Mejorar la accesibilidad y usabilidad para atraer más usuarios
2. Vincular los conjuntos de datos relacionados entre sí
3. Relacionar datos abiertos con datos geoespaciales
4. Mejorar la administración y uso de metadatos
5. Disminuir el costo y el trabajo necesario para publicar los datos
6. Introducir el historial de revisiones
7. Mejorar la gestión de grandes conjuntos de datos

2.3. Transparencia

Cuando un objeto nos permite observar con nitidez lo que hay dentro o detrás de él, decimos que el objeto es transparente. Por ejemplo un vaso de vidrio que nos permite ver el contenido que hay en su interior. Así la idea de transparencia en el gobierno busca permitir que los ciudadanos conozcan el proceder del gobierno y evaluar si su trabajo se está realizando de manera justa y equitativa.

Hay una organización internacional denominada *Transparencia Internacional* encargada de construir y publicar un Índice de Percepciones de la Corrupción (IPC). Para ellos la transparencia es “la cualidad de un gobierno, empresa, organización o persona de ser abierta en la divulgación de información, normas, planes, procesos y acciones. Como regla general, los funcionarios públicos, empleados públicos, gerentes y directores de empresas y organizaciones, y las juntas directivas tienen la obligación de actuar de manera visible, predecible y comprensible en la promoción de la participación y la rendición de cuentas”. Citado por Cruz y compañía [28]

Si bien es cierto que la transparencia y la rendición de cuentas son una herramienta importante contra la corrupción, no se debe considerar solo como un mecanismo para disminuir este mal. Lo que para un gobierno puede significar un deber tedioso, costoso e innecesario para otro puede ser un aliado para gobernar sabiamente, pues si un ciudadano está informado de los asuntos públicos puede interesarse y comprometerse con el desarrollo social y político de su país. O sea si el gobierno es transparente y ofrece los datos de su gestión para que los ciudadanos puedan participar, opinar y tomar buenas decisiones, se mejora la gobernanza. Según un estudio realizado por el Pew Research Institute titulado “How the public perceives community innovation systems” citado por Concha y Naser [26] entre más abierto es un gobierno municipal hay mayor participación ciudadana y se generan mejores prácticas gubernamentales que desde el momento en que se crean cuentan con la aprobación de los ciudadanos.

La figura 2.4 ayuda al lector a comprender que existen organizaciones internacionales que los países en estudio integran o están intentando ingresar y que estas organizaciones exigen a sus integrantes legislación sobre transparencia y medidas anticorrupción. Algunas de estas organizaciones pretenden medir el grado o percepción de corrupción, y estos índices cobran importancia pues demuestran falencias y aspectos por mejorar en sus evaluados. Por otro lado, estas debilidades se convierten en barreras para el ingreso a las organizaciones y terminan forzando a los países a adoptar convenciones, crear reglamentos y promulgar leyes y decretos afines.

Aunque los datos abiertos y la transparencia ofrecen grandes oportunidades y beneficios, también contienen grandes retos que se deben legislar y manejar con cuidado. Uno de ellos es la protección de los datos personales, o sea, aquellos datos que permitan identificar o hacer identificable a una persona, por ejemplo, el nombre, huella digital, número de cédula, fotografía o incluso hasta ADN. Como se mencionó anteriormente los países han creado leyes en este sentido pero algunas de ellas fueron creadas en el siglo pasado, lo que conlleva una desactualización con respecto al panorama actual en el que los dispositivos móviles, las empresas de salud y otras muchas entidades gestionan grandes cantidades de datos personales.

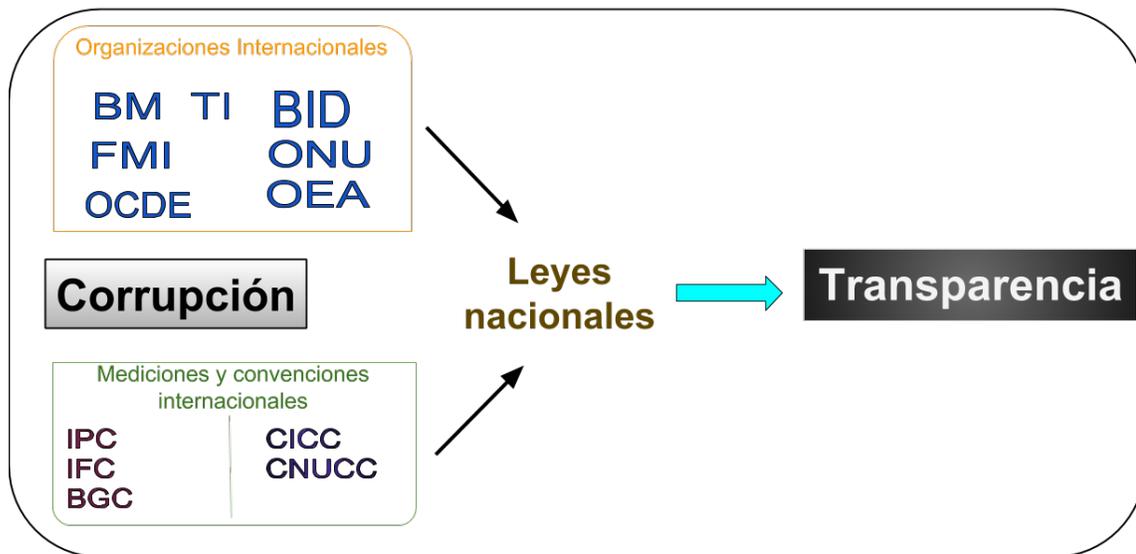


Figura 2.4: Flujo hacia la transparencia
Fuente: Creación propia.

Por ejemplo, según plantea Viollier [98], en Chile “la ley actual parece definir correctamente conceptos básicos en cuanto al tratamiento de datos personales. Define lo que debe entenderse por datos personales, datos sensibles, datos caducos y lo que es un banco de datos, sin embargo falla al intentar una definición precisa de fuente de datos accesibles al público, con lo cual, en la práctica queda desprotegido el tratamiento de datos personales en el país” .

Los representantes de la sociedad civil al enterarse que los gobiernos publican datos abiertos, realizan eventos para conocer las problemáticas sociales que pueden ser resueltas usando las técnicas de análisis científico con los datos liberados. De esta forma surgieron los *hackatones* -viene de la unión de hacker y marathon- en donde equipos de programadores, diseñadores y periodistas se reúnen para analizar y resolver problemas sociales a través de la tecnología. Al sector social que desarrolla este tipo de trabajos de bien social se le conoce como hacker cívicos. En su trabajo Caffarena [14] pretende dar a entender qué impacto tienen las políticas de transparencia de las instituciones públicas en el desarrollo de esta práctica, llamada también activismo de datos. En el trabajo denominado *Soluciones tecnológicas a problemáticas sociales y ciudadanas. El caso de los Retos Cívicos y Públicos en México* [4] se describen las iniciativas, desafíos cívicos y retos públicos en dicho país; así como las experiencias de innovación en la implementación de estrategias tecnológicas. Como se mencionó anteriormente, en Montevideo se realizaron una serie de aplicaciones para facilitar al ciudadano la utilización de servicios público como el transporte y para ubicar los puntos de reciclaje en la ciudad, entre otros. Por ejemplo, conociendo los datos de los centros educativos, los ciudadanos pueden elegir con mayor criterio en cual inscribir sus hijos, o consultando en una aplicación donde se encuentran las principales fallas tectónicas se tiene criterio sobre donde conviene no comprar un terreno. Esta posibilidad de crear valor a través de los datos abiertos genera una mayor fuerza hacia el gobiernos por parte de la sociedad civil para que éste entregue sus datos de forma que sean fácilmente reutilizables.

La figura 2.5 ayuda comprender al lector que la transparencia surge en parte como un mecanismo contra la corrupción, pero también como una característica del gobierno demo-

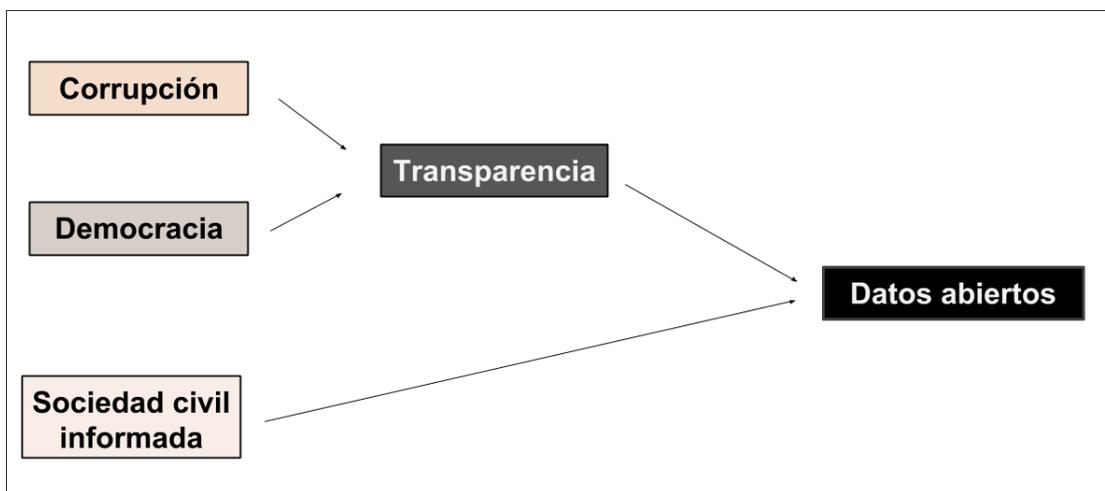


Figura 2.5: Interacción entre los factores que inciden en la publicación de datos abiertos
Fuente: Creación propia.

crático. En un gobierno democrático el ciudadano participa activamente y presiona para que las autoridades rindan cuentas y muestren libremente los datos de su gestión. Como se mencionó anteriormente, a partir de los datos que el gobierno libera se puede crear valor, estas iniciativas de la sociedad civil que buscan utilizar los datos para construir bienestar social, para procesarlos y generar información que ayude en la correcta toma de decisiones son cada día son mayores, por ello presionan al gobierno para que liberen *datos abiertos*.

Los datos digitales son menos costosos y tienen menos limitaciones para ser obtenidos y procesados que los documentos físicos. Con la ayuda de la legislación de gobierno electrónico es posible entregar datos digitales con la misma validez legal que los físicos. Según datos de la ONU [70] para el 2016 Uruguay ocupó el lugar número 34 en la encuesta de Gobierno Electrónico; Chile que en el 2014 ocupó el puesto 33 y descendió 9 puestos ocupando la casilla 42; Costa Rica por su parte asciende un puesto y aparece en la posición 53.

El Gobierno Abierto facilita la rendición de cuentas y la transparencia, lo que a su vez promueve la participación ciudadana, y permite procesar datos digitales que ayudan a mejorar las condiciones de vida y apoyan la toma de decisiones tanto de los ciudadanos como del mismo gobierno.

2.3.1. Transparencia activa y pasiva

Cuando un gobierno se compromete a transparentar su gestión permite que los ciudadanos puedan solicitar la información que requieran a través de una ventanilla o de un medio digital, sin necesidad de justificar su solicitud. En este caso el Estado como ente pasivo no construye información para el usuario, ni publica, sino solamente brinda el acceso a la información según se demande. A este tipo de transparencia se le llama *transparencia pasiva*.

Por el contrario, *la transparencia activa* se manifiesta cuando el Estado garante de la transparencia se esfuerza por conocer qué información le interesa al pueblo y prepara dicha

información para ser publicada digital o físicamente. En la actualidad debido a la gran cantidad de datos que poseen los gobiernos, es difícil proveer datos en forma física y a través de documentos. Por ello, como se mencionó en la sección de datos abiertos los gobiernos se han visto en la necesidad de brindar los datos por medio de API's, o en los formatos que se detallaron, de manera que se puedan manipular por medio de computadores.

2.3.2. Transparencia y acceso a la información

En el año 1766 se dictó en Suecia la Real Ordenanza sobre la Libertad de Prensa que declaraba el derecho a acceder a documentos oficiales. Este primer hito fue precedido por Ley sobre la Naturaleza Pública de los Documentos Oficiales, dictada en en Finlandia en 1951 y por la Ley de Libertad de Información promulgada en 1966 en Estado Unidos. Posteriormente se dictaron varias leyes con el mismo espíritu alrededor del mundo.

Para el caso de América Latina fue en 1985 que se dio a conocer la ley 57 que ordena la publicidad de actos y documentos oficiales en Colombia. En Chile y Uruguay la ley sobre Acceso a la Información Pública se creó en el 2008, mientras que Costa Rica todavía no cuenta con una ley al respecto, al parecer el expediente No. 16.198 es el proyecto de ley, pero aún no ha sido aprobado.

En marzo del 2017 se firmó en Costa Rica la carta de entendimiento de Gobierno Abierto. Posteriormente, el ejecutivo emitió el decreto No. 40199-MP “Apertura de datos públicos” y la directriz ejecutiva No. 074-MP “sobre la apertura de datos abiertos”, con el objetivo definir los lineamientos que deberán cumplir las instituciones para lograr la apertura de los datos.

Estas leyes y decretos buscan que las instituciones públicas pongan a disposición de la ciudadanía en forma clara y abierta sus datos, a excepción de los documentos que se consideran secreto de estado. Además, permiten a los ciudadanos conocer las actividades y decisiones que toman sus gobiernos. En este sentido, en vez de que solo haya acceso a la información se busca que los gobiernos transparentes su gestión, el uso de recursos tanto materiales como financieros.

2.3.3. Transparencia y corrupción

En la actualidad, la mayoría de los países de América Latina poseen gobiernos democráticos, los cuales presumen de la igualdad de los derechos y obligaciones de los ciudadanos, controles institucionales, libertad de expresión, poder judicial autónomo, sociedad civil organizada e informada. Sin embargo, muchas veces la calidad de la democracia se ve opacada por los cuestionamientos en las concesiones y acusaciones de enriquecimiento ilícito, entre otras cosas. En este contexto, la transparencia se plantea como un medio para ayudar a evitar estos problemas.

Según Cruz y compañía [28] “Es precisamente en la agenda de la calidad de la democracia donde se localiza la cuestión de la corrupción”. El trabajo de Cruz et al. ofrece un recorrido

sobre las causas y consecuencias de la corrupción, su conexión con el desarrollo y algunas medidas que se pueden implementar para contrarrestarla. A continuación se aborda el tema de corrupción considerando su concepto y algunas medidas que se han tomado alrededor del mundo para mitigarla.

Para la convención Interamericana contra la corrupción, la corrupción constituye:

El requerimiento o aceptación; el ofrecimiento u otorgamiento, directos o indirectos, por o a un funcionario público o persona que ejerza funciones públicas, de cualquier objeto de valor pecuniario u otros beneficios como dádivas, favores, promesas o ventajas para sí mismo o para otra persona o entidad a cambio de la realización u omisión de cualquier acto en el ejercicio de sus funciones públicas; la realización por parte de un funcionario público o una persona que ejerza funciones públicas de cualquier acto u omisión en el ejercicio de sus funciones, con el fin de obtener ilícitamente beneficios para sí mismo o para un tercero; el aprovechamiento doloso u ocultación de bienes provenientes de cualesquiera de los autor, instigador, cómplice, encubridor o en cualquier otra forma en la comisión, tentativa de comisión, asociación o confabulación para la comisión de cualquiera de los actos mencionados. Citado por Cruz y compañía [28]

Pasquino citado por Cruz et al. [28] define corrupción como: “una transacción o un intercambio entre quien corrompe y quien se deja corromper. Normalmente se trata de la promesa de una recompensa a cambio de un comportamiento que favorezca los intereses del corruptor; raramente se amenaza con castigos en caso de comportamiento que perjudique los intereses del corruptor. Este intercambio negativo es mejor definido como coerción”

Dado el alto grado de corrupción internacional, al pago de sobornos por parte de empresas extranjeras para que les adjudicaran grandes trabajos, se escapaba de las manos de las autoridades judiciales las posibilidades de dar seguimiento y castigar a los corruptos. Por ello, organizaciones como Transparencia Internacional, fundada en 1993, realizaron esfuerzos para medir la corrupción a través de índices como: el Índice de Percepciones de la Corrupción (IPC) en 1995, el Índice de Fuentes de Soborno (IFS) en 1999 y el Barómetro Global de la Corrupción (BGC) en 2003.

Las organizaciones internacionales que tienen entre sus objetivos mejorar el desarrollo de los países y financiar ayudas en un contexto internacional, tales como: la Organización para la Cooperación y el Desarrollo Económicos (OCDE) el Banco Interamericano de Desarrollo (BID), Banco Mundial (BM), Fondo Monetario Internacional(FMI) también realizan encuestas de diagnóstico a nivel local, regional, sectorial, para contribuir a la lucha contra la corrupción, verificar que sus aportes estén siendo dirigidos a lugares que no sean corruptos y mejorar el bienestar de las poblaciones.

Con este panorama, se han creado una serie de convenciones e instrumentos legales internacionales entre ellos, los acuerdos tomados en la cumbre de las Américas de Miami, y en la conferencia internacional sobre ética pública realizadas en 1994. Dos años después, en 1996, la Organización de Estados Americanos (OEA) celebró la Convención Interamericana contra la Corrupción (CICC), posteriormente la Convención de la OCDE contra el soborno transnacional, en vigor desde 1999, y finalmente la Convención de las Naciones Unidas contra

la Corrupción (CNUCC) en el 2003.

2.3.4. Principales hitos de transparencia en Chile, Costa Rica y Uruguay

A continuación se describen los hitos más importantes relacionados con transparencia acontecidos en los países estudiados. La tabla 2.2 resume para el lector dichos hitos.

Chile El gobierno de Chile en octubre de 1998, a través de su representante permanente en la Organización de los Estados Americanos ratifica la Convención Interamericana contra la Corrupción. Posteriormente en 1999 se publica la ley 19.628 sobre la protección de la vida privada o ley de protección de datos de carácter personal como se le llama comúnmente y en el año 2012 la ley 20.575 sobre el tratamiento de datos personales. En relación a estas leyes, en marzo del 2017 la Jefa de Estado chilena firmó un proyecto de ley que regula la protección y el tratamiento de los datos personales. También en 1999 se aprobó ley 19.653 sobre probidad administrativa aplicable de los órganos de la administración del Estado, a la que podríamos llamar ley anticorrupción. Como se mencionó anteriormente dado que las acciones de corrupción traspasan las fronteras de los países, se hizo necesaria la coordinación y el compromiso por medio de convenciones, en este sentido en el año 2002 a través del decreto 496 se promulga la convención para combatir el cohecho a funcionarios públicos extranjeros en transacciones comerciales internacionales de la Organización para la Cooperación y el Desarrollo Económico (OCDE). En un sentido similar, sobre el cohecho a un funcionario público nacional y extranjero y sobre el delito cometido por una persona natural en beneficio de alguna empresa se publica la ley 20.393. Dicha ley se dictó en 1999, bajo el título “establece la responsabilidad penal de las personas jurídicas en los delitos de lavado de activos financiamiento del terrorismo y delitos de cohecho que indica” como parte de los requisitos que necesitaba el país para ingresar a la OCDE. El 23 de noviembre de 2006 se adopta la Convención de las Naciones Unidas contra la Corrupción, y como se mencionó anteriormente en el año 2008 se crea la Ley 20.285 sobre acceso a la información pública. En la página Web [42] de gobierno digital hay una recopilación de los decretos supremos que se vinculan con la administración de los sitios Web del Estado y de las normas técnicas sobre documentos y comunicación electrónica. Cira Alvear [5] elaboró un compendio de gestión pública que incluye un apartado de normas relativas a la información, probidad y transparencia. La página anticorrupcion.cl [6] ofrece una lista de de las leyes nacionales, decretos y principales convenciones internacionales anticorrupción.

En octubre de 2012 Chile toma la iniciativa de adherirse a la Alianza de Gobierno Abierto (Open Government Partnership, OGP), luego de haber realizado una reforma constitucional en materia de transparencia, modernización del estado y calidad de la política (2010), mediante la ley 20.414. En su camino al ingreso a la OGP el gobierno de Chile realizó una serie de reuniones con las organizaciones civiles y organizaciones internacionales para construir los planes de acción.

Costa Rica Respecto a la legislación local de Costa Rica, este país consagra en el artículo 30 de su Constitución política el derecho fundamental de acceso a la información pública de la siguiente manera “Se garantiza el libre acceso a los departamentos administrativos con propósitos de información sobre asuntos de interés público. Quedan a salvo los secretos de Estado que Costa Rica considere necesario”. La Convención Interamericana contra la Corrupción de la OEA fue adoptada en 1997. Posteriormente se promulga la ley contra la corrupción y el enriquecimiento ilícito en la función pública en el año 2004, y también se ha modificado el código penal para castigar la corrupción, (para más detalles de los artículos modificados puede consultar [18] en la página 229). Tres años después de que las Naciones Unidas celebraron el Convenio contra la Corrupción, o sea, en el año 2006, este país dio su adhesión. Aunque el acceso a la información pública es un derecho constitucional, es necesario regular la publicación para proteger los derechos de privacidad de los datos personales. En este sentido para el año 2011 los costarricenses publican la ley de Protección de la persona frente al tratamiento de sus datos personales.

En relación al Gobierno Abierto es hasta el año 2012 que el país inicia su actividad formando parte de la OGP y creando en el siguiente año el primer plan de acción y un portal web más orientado hacia el gobierno electrónico que hacia el Gobierno Abierto, pero con una visión futura a la transparencia. Para el año 2014, el gobierno y la sociedad civil promueven la inclusión del Gobierno Abierto en el plan nacional de desarrollo, por lo que el gobierno incluye como tercer pilar del plan el Gobierno Abierto y la transparencia. Más recientemente el país ha trabajado fuertemente en el tema. Desde mayo del 2015 las organizaciones civiles, otras entidades y el gobierno trabajaron en conjunto para definir el segundo plan de acción, que fue presentado en octubre del mismo año. El último hito importante en ese año fue que el ejecutivo declaró de interés público la Estrategia Nacional por un Gobierno Abierto 2015-2018. De los tres países estudiados, Costa Rica es el que adoptó más recientemente el modelo de estado abierto, sin embargo, ha avanzado mucho, pues se le reconoce la iniciativa en los tres poderes, esto quiere decir que la apertura se ha extendido al plano legislativo y judicial, cada uno con su plan de acción.

Uruguay En Uruguay, encabeza la lista de las normativas legislativas la ley N° 17.060 decretada en 1998, conocida como ley cristal o ley anticorrupción, seguida por la ley 17.008 que contiene la aprobación de la Convención Interamericana contra la corrupción en ese mismo año. Para el año 2005 se crea la Agencia de Gobierno electrónico y Sociedad de la Información y del Conocimiento (Agesic), con el objetivo de que esta lidere la estrategia de gobierno electrónico y su implementación en el país. En ese momento Agesic no estaba vinculada a los datos abiertos, pero posteriormente contribuyó en gran manera. La legislación relativa a la aplicación de la Convención de las Naciones Unidas contra la Corrupción es la Ley 18.056, la cual fue aprobada por el Senado y la Cámara de representantes el 14 de noviembre de 2006; por medio de esta ley, la Convención pasó a formar parte del ordenamiento jurídico de Uruguay. Como se mencionó anteriormente, en el año 2008 se creó la Ley N° 18.381 que regula el Derecho de acceso a la información pública, y con ella el mandato de la creación de la Unidad de acceso a la información pública (UAIP). Este es un órgano especializado cuyo cometido es velar por el respeto al derecho de acceso a la información pública. En agosto del mismo año se dictó la Ley N° 18.331 conocida como la ley de Protección de datos personales y acción de habeas data, en la cual se incluye la creación de la Unidad Reguladora y de

Control de Datos Personales (URCDP) que custodia el cumplimiento de la legislación de protección de datos personales y asegura el respeto de sus principios. También en el año 2008 se creó la Agenda Digital de Uruguay (ADU) con el fin de mantener alineada la estrategia de Gobierno Electrónico con las últimas tendencias y mejores prácticas mundiales, fue en esta misma agenda en su versión 2011-2015 donde se promulga la iniciativa de datos abiertos en ese país. En el año 2010 durante la construcción de la ADU se crea el compromiso de abrir los datos y como primera acción se crea el portal de datos abiertos. En el año 2011 se creó el primer plan de acción, hito importante relacionado con su adhesión a Alianza de Gobierno Abierto en inglés Open Government Partnership (OGP), realizada en ese mismo año.

Para lograr tener datos abiertos en todas las instancias públicas de forma estándar en Uruguay, desde el año 2012 se elaboran y publican guías y se realizan talleres de publicación de datos para los funcionarios públicos. Una vez con datos abiertos se crean concursos como *Dataidea* que promueven el uso de los datos a través de aplicaciones o ideas de servicios a los ciudadanos a través de datos. También se ha puesto a disposición un sitio para crear, gestionar y publicar colecciones de conjuntos de datos, usando CKAN que es una herramienta de código abierta desarrollada por la Fundación de Conocimiento Abierto en inglés Open Knowledge Foundation (OKF).

Para fomentar en la ciudadanía uruguaya la utilización y reutilización de datos abiertos, en el año 2013 se realizan talleres a periodistas y medios de prensa en general, además talleres formativos para estudiantes de comunicación e informática. En ese mismo año los uruguayos realizan la primera conferencia regional de datos abiertos. Dos hechos destacables durante el año 2014 fueron: primero, la celebración del primer Datacamp (realizado también en el 2015 y 2016) en el cual se integran periodistas, desarrolladores y diseñadores para aprender cómo encontrar, extraer y analizar datos públicos. Segundo, la celebración del concurso Dataidea, en el que se destacaron actividades como: jornadas de análisis de problemas a resolver mediante servicios públicos que utilicen las tecnologías de la información y jornadas de desarrollo y creación de prototipos relacionados a los problemas analizados. En octubre del 2015 Uruguay se adhiere a la Carta Internacional de Datos Abiertos. Bertón presenta un trabajo [11] sobre la participación ciudadana en Uruguay a partir de las opiniones de referentes académicos en la materia, los encuestados exponen su opinión sobre la participación ciudadana electrónica. En la compilación realizada por Patiño [78], Fabrizio Scrollini cita una serie de aplicaciones creadas en Uruguay que podrán orientar al lector sobre usos digitales reales en la región utilizando datos abiertos. Entre los nombres de aplicaciones que se destacan están: ¿Dónde reciclo?, Sistema de Información Geográfica (MonteviMap), Montevideo Bus.

La tabla 2.2 resume los hitos más importantes de los desarrollos comparados de Chile, Costa Rica y Uruguay. En el caso de Chile y Costa Rica existen iniciativas constitucionales que sientan las bases de la libertad de acceso a la información. También se puede observar que los países en estudio se han adherido a convenciones internacionales a favor de la transparencia, quizá como primer paso para mitigar los sobornos por parte de las transnacionales hacia los políticos con poder de decisión. En el contexto local, cada país ha desarrollado un cuerpo jurídico hacia la transparencia, que incluyen leyes con la intención de disminuir la corrupción a nivel nacional, leyes de protección de datos, leyes de transparencia (en proyecto de ley en Costa Rica), leyes de participación ciudadana (ausente en Costa Rica) y algunas leyes y decretos ejecutivos que ponen en evidencia la iniciativa de actualizar tecnológicamente el

Estado, por ejemplo, la legislación sobre documentos electrónicos y firma digital.

Evento	Chile	Costa Rica	Uruguay
Referencia constitucional	2005. Ley 20050 Modifica la constitución obligando a que “toda función pública se realice actuando según el principio de probidad, e indicando que son públicos los actos y resoluciones de los órganos del Estado, así como sus fundamentos y los procedimientos que utilice”	2000. “La Administración Pública en sentido amplio, estará sometida a un procedimiento de evaluación de resultados y rendición de cuentas...”. Modificado por la Ley N° 8003” y el Artículo 30 “garantiza el libre acceso a los departamentos administrativos con propósitos de información sobre asuntos de interés público. Quedan a salvo los secretos de Estado...”.	
Convención interamericana contra la corrupción	1999. Mediante Decreto 1879 y por oficio N°2.145 de la Cámara de Diputados.	1997. Ley 7670 Asamblea legislativa	1998. Mediante ley N.° 17008
Convención de las Naciones Unidas contra la corrupción	2006. Decreto 375 y en el oficio N° 6223 de la Cámara de Diputados.	2006. Mediante ley N° 8557 de la asamblea legislativa	2006. Mediante ley N.° 18056
Convención Anticohecho de la OCDE	2002. Decreto supremo 496	2017. Ley 9450 Asamblea legislativa	
Continúa en la siguiente página			

Continuación de la tabla 2.2

Evento	Chile	Costa Rica	Uruguay
Protección de datos personales	1999. Ley 19628 sobre protección de la vida privada 2012. Ley 20575 establece el principio de finalidad en el tratamiento de datos personales	2011. Ley 8968 de protección de la persona frente al tratamiento de sus datos personales	2008. ley 18331 sobre la Protección de datos personales y acción de "habeas Data". 2012. Ley N° 19.030 adopta el convenio para la protección de las personas con respecto al tratamiento automatizado de datos de carácter personal.
Ley Anticorrupción	1999. Ley 19653 sobre probidad administrativa aplicable de los órganos de la administración del estado.	2004. Ley 8422 contra la Corrupción y el Enriquecimiento Ilícito en la Función Pública 2016. Ley 20880 sobre probidad en la función pública y prevención de los conflictos de intereses	1998. Ley 17.060 conocida como ley cristal o ley anticorrupción
Ley de transparencia	1990. Decreto 680 aprueba instrucciones para el establecimiento de oficinas de información para el público usuario en la administración del estado 2006. Instructivo presidencial sobre transparencia activa y publicidad de la información de la administración del Estado. 2008. Ley 20.285 Sobre acceso a la información pública	1990. Ley 7202 del Sistema Nacional de Archivos. Garantiza el libre acceso a todos los documentos que produzcan o custodien las instituciones, de los poderes Legislativo, Judicial y Ejecutivo, y de los demás entes públicos. 2017. El expediente 16.198 es el proyecto de ley, pero aún no ha sido aprobado.	2008. Ley 18.381 de acceso a la información pública
Ley de participación ciudadana	2011. Ley 20.500 sobre asociaciones y participación ciudadana en la gestión pública		2009. Ley N° 18.567 descentralización política y participación ciudadana

Continúa en la siguiente página

Continuación de la tabla 2.2

Evento	Chile	Costa Rica	Uruguay
Gobierno digital	<p>2002. Ley 19799 sobre documentos electrónicos, firma electrónica y servicios de certificación de dicha firma</p> <p>2004. Decreto 77 aprueba norma técnica sobre eficiencia de las comunicaciones electrónicas entre órganos de la administración del estado y entre estos y los ciudadanos (Derogado 2014)</p> <p>2004. Decreto 81 aprueba norma técnica para los órganos de la administración del estado sobre interoperabilidad de documentos electrónicos (Derogado 2014)</p>	<p>1990. Ley N° 7169 de Promoción del Desarrollo Científico y Tecnológico, que Impulsa la incorporación de tecnología en la administración pública, a fin de agilizar y actualizar, permanentemente, los servicios públicos, para lograr la modernización del aparato estatal</p> <p>2005. Ley 8454 de certificados, firmas digitales y documentos electrónicos.</p> <p>2008. Ley 8660 fortalecimiento y modernización de las entidades públicas del sector telecomunicaciones, entre sus objetivos está garantizar la rendición de cuentas y la evaluación de resultados por parte del Instituto costarricense de electricidad y sus empresas.</p>	<p>1996. Ley 16736 en sus artículos 694, 695, 696 sobre el empleo y aplicación de medios informáticos y telemáticos para el desarrollo de las actividades y el ejercicio de las competencias de la administración pública.</p> <p>1998. Decreto N° 65/998 sobre expediente electrónico (Derogado en 2013)</p> <p>2003. Decreto N° 382/003 reglamentación del uso de la firma digital y su reconocimiento jurídico (Derogado 2013)</p> <p>2009. Ley N° 18600 documento electrónico y firma electrónica. admisibilidad, validez y eficacia.</p> <p>2013. Ley N° 19.179 sobre la utilización de software libre y formatos abiertos en el aparato estatal.</p>
Continúa en la siguiente página			

Continuación de la tabla 2.2			
Evento	Chile	Costa Rica	Uruguay
Alianza para el Gobierno Abierto	2011. Ingreso a la OGP 2015. Adopta la Carta Internacional de Datos Abiertos	2012. Ingreso a la OGP 2013. Creación del primer plan de acción 2013-2015 2014. El Gobierno Abierto y transparencia es uno de los 3 pilares del Plan Nacional de Desarrollo, 2015-2018 2015. Octubre, se publica el segundo Plan de Acción de Gobierno Abierto 2015-2017 2016. Adopta la Carta Internacional de Datos Abiertos	2011. Ingreso a la OGP 2015. Adopta la Carta Internacional de Datos Abiertos 2015. El artículo 82 de la ley 19.355 dicta la publicación de la información, datos y sus metadatos (como mínimo publicar en formato abierto) asociados según las normas técnicas que determine la Agesic.

Tabla 2.2: Resumen de los hitos hacia la transparencia y los datos abiertos en Chile, Costa Rica y Uruguay

Fuente: Creación propia

2.3.5. Trabajos relacionados sobre evaluación de la transparencia

El trabajo presentado por Fierro y compañía [38] es un diagnóstico general del estado actual de la política de transparencia y del acceso a la información pública en los tres poderes de gobierno de los tres niveles de gobierno y en algunos órganos constitucionales autónomos. La primera edición se realizó en el 2010 y su más reciente versión se publicó en el 2014, ambos en México. Para el 2014 se actualizaron sus métricas y metodologías. El aporte de este trabajo para la propuesta de tesis a desarrollar es la posibilidad de estudiar sus métricas en especial porque el estudio posee una dimensión que evalúa los portales de transparencia.

Los trabajos desarrollados por Zamora [102] y Seveso y compañía [90] son índices que permiten evaluar los sitios de transparencia, el primero desarrollado y aplicado en Costa Rica y el segundo en Uruguay. Por su parte [64], es parte de un par de trabajos compuestos por una guía de diseño para portales web transparentes y una base metodológica para la guía, ambos trabajos aportan para la tesis a desarrollar, aunque el mayor aporte está en la base metodológica, pues propone métricas e indican cómo elaboran la guía, esta información nos puede orientar en el trabajo de tesis por realizar.

Schiavi [88] realiza un análisis de lo que él propone como las claves para el éxito de la gestión de los gobiernos municipales, cita la participación ciudadana, la transparencia y la articulación entre los distintos niveles de administración (local, regional, nacional). Analiza las obligaciones que tienen los gobiernos a la luz de la ley N° 18381 sobre Acceso a la información pública, obligaciones de transparencia activa y pasiva. Este trabajo nos permite tener un acercamiento a la legislación y definición de términos referente en Uruguay, y que posteriormente se comparan con los de otros países de interés en la región.

Considerando la experiencia del gobierno de los EEUU, Soto [91] plantea una serie de desafíos que tendrán que enfrentar los gobiernos en la búsqueda de alcanzar la transparencia. Por otro lado, Cruz et al. [28] desarrolla el concepto de corrupción, explora su causas, costos y consecuencias; posteriormente enuncia las razones por las cuales considera que las soluciones al problema residen, en la activación de mecanismos de transparencia y rendición de cuentas.

Cejudo [17] presenta un trabajo en el que estudia 4 países de la misma región sobre la que se realizará esta propuesta de tesis, en el cual muestra una comparación entre los resultados de un examen sobre el acceso a los portales de información de transparencia de Honduras, Nicaragua, Panamá y Guatemala. Es importante destacar que dicho trabajo realiza una comparación en términos de información y que no se acerca a los datos. Toma como base una metodología Fierro [38] creada en México de la cual se hablará en el siguiente apartado.

Carrasco [16] muestra las dificultades que puede presentar la información de transparencia desde un punto de vista más orientado a la comunicación, e indica que la “apertura de datos bajo la iniciativa de gobierno electrónico no es suficiente para ganar la confianza de los ciudadanos y que lo que se requiere es construir un verdadero modelo de interacción entre los ciudadanos y gobierno”.

Para tener sustento en lo referente a las leyes que se relacionan con acceso a la información pública y transparencia en Costa Rica, se consultará el “Compendio de leyes costarricenses sobre acceso a la información pública y transparencia” [27] un trabajo publicado por el Dr. Córdoba en el 2013. Por supuesto considerando los últimos avances sobre el tema como la firma del convenio de estado abierto, que es un marco para promover un estado abierto de la república de Costa Rica entre el poder ejecutivo, el poder legislativo, el poder judicial y el tribunal supremo de elecciones.

Capítulo 3

Producción, usos, usuarios y dimensiones de los datos abiertos

Es posible que si evaluamos el desempeño de un martillo para colocar un tornillo en una pared los resultados sean negativos, considerando que el martillo no es un instrumento para atornillar. Por ello, antes de evaluar un objeto es necesario tener claro cuál es el uso del objeto. Para el caso de los datos abiertos es importante conocer qué tipo de usuarios utilizan los datos abiertos y cuáles son los usos más comunes. En este capítulo se estudian los usos más comunes de los datos abiertos, y se propone una clasificación de sus usuarios, con miras a construir un marco para evaluarlos. Luego de estudiar las características de los datos y el contexto en que se utilizan los datos abiertos se propone un conjunto de dimensiones que permitan evaluar y clasificar los datos publicados en los sitios de gobierno, con vista a diseñar los indicadores que permitirán realizar la evaluación. Se espera que la propuesta metodológica permita reconocer a quien están dirigidos estos sitios de gobierno.

También se describen dos formas de obtener datos. La primera es captando los datos de una *fente primaria*, por ejemplo del ambiente. La segunda es creando datos a partir de otros datos, *fente secundaria*. También se explican dos mecanismos de producción de datos que dependen de la regularidad con la cual se producen tomándolos de una misma fuente, a estos mecanismos se les ha denominado *producción de datos estática* y *producción de datos dinámica*.

3.1. Usos y usuarios de los datos abiertos

Una revisión de la literatura reciente realizada por Safarov et al. [85] indica que los principales usuarios de los datos abiertos son: ciudadanos, comerciantes, investigadores, desarrolladores, organizaciones no gubernamentales y periodistas. Y que los principales usos son: innovación, análisis de datos, toma de decisiones, anticorrupción, ciudades inteligentes, investigación, nuevos servicios y por últimos hackatones y competiciones.

Uno de los principales motivos por los cuales los gobiernos liberan datos es para forta-

lecer los mecanismos de transparencia, permitiendo así que los ciudadanos puedan evaluar el trabajo que los gobiernos realizan, además de participar más activamente en el diseño de políticas y en la toma de decisiones. Por tanto, los *ciudadanos* son sin duda alguna usuarios de los datos de gobierno. ¿Qué características pueden tener los datos que se le brindan a los ciudadanos? Cuando se muestra un cuadro de datos es importante colocar una descripción para que el destinatario comprenda de qué le están hablando; es relevante conocer quién publica los datos; y de ser posible tener los datos suficiente para trazar su origen; por tanto, la documentación del conjunto de datos juega un importante papel. Los datos deben estar completos y ser estandarizados, coherentes. Si se publican varios conjuntos de datos es bueno tener un catálogo de datos que permita buscar con facilidad, entre otras cosas.

Por otro lado, es evidente que la cantidad de datos que generan los gobiernos a través de sus oficinas no es poca, en la mayoría de los casos más que megabytes, quizás unos gigabytes o incluso terabytes. El caso es que manejar estos datos en forma manual se vuelve imposible, en especial si se pretende aprovechar el valor de los datos históricos publicados los cuales en conjunto fácilmente pueden alcanzar unos gigabytes. Algunas empresas conociendo la cantidad de datos publicados y el valor que puede adquirir lo que se produzca con esos datos han decidido hacer uso de ellos para incrementar sus ganancias.

En estas empresas se pueden observar algunos usuarios. Por ejemplo, los que aportan capital, adquieren equipos y software, que también se preocupan de las reglas del negocio, es decir, que los datos tengan licencia abierta, que se publiquen el formato abierto para no tener que comprar software costoso. Estos manejan la estrategia del negocio y definen qué productos van a ofrecer, por tanto necesitan conocer si los datos disponibles están actualizados o si hay históricos, que los datos sean pertinentes de acuerdo al producto que intentan comerciar. Estos usuarios son *gerenciales*.

Con otra perspectiva, pero dentro de la misma empresa están las personas que trabajan con los datos, que les interesa conocer detalles de las API, de seguridad, existencia de metadatos, interoperabilidad, entre otros. Estos son usuarios *desarrolladores*.

Con un enfoque más analítico y más crítico que los ciudadanos se encuentran quienes usan los datos para investigar, ya sea para crear un artículo periodístico o académico. Poseen un perfil profesional y herramientas analíticas avanzadas, este grupo involucra a periodistas e investigadores. Para este trabajo se les resumió con el nombre de *investigadores*.

Considerando las actividades que desarrollan estos grupos de usuarios y sus características ampliamos los horizontes y realizamos una primera clasificación que derivó los siguientes subgrupos:

1. **Ciudadanos:** Comprende los individuos interesados en conocer la labor que realizan sus gobiernos y que desean contribuir con el bienestar social y con la gobernanza. Además del colectivo o grupos de individuos que se organizan con los fines antes mencionados.
2. **Desarrolladores:** Esta categoría se subdivide en tres: primero, los desarrolladores cívicos aquellos que no realizan la labor por una paga si no por el bienestar común, participan en hackatones o simplemente desarrollan como parte de sus pasatiempos. Segundo, los desarrolladores comerciales, aquellos que crean productos o servicios con

fines lucrativos. Tercero, los desarrolladores de las instituciones y organismos públicos.

3. **Gerenciales:** Aquellas personas que administran capital público o privado con poder de toma de decisiones, que utilizan los datos públicos de forma estratégica para mejorar políticas, o con fines lucrativos.
4. **Investigadores:** Comprenden tanto a los periodistas que trabajan para crear productos de información para el consumo público como también a los investigadores académicos que posiblemente combinan datos de diversas fuentes para probar y confirmar nuevas hipótesis.

Por otro lado, el análisis de esta primera clasificación y el estudio de la analogía del ecosistema de datos abiertos permitió dividir a los usuarios en dos grandes grupos: Los que utilizan los datos para crear otros datos o productos de información denominados *productores* y los que consumen los datos solamente, conocidos como *consumidores*.

A continuación se muestran algunos escenarios de creación de datos e información. En general los datos se pueden obtener a través de sensores o producirse a partir de otros datos. Por ejemplo, en la figura 3.1 se ilustra este primer caso en el que un sensor de temperatura capta de forma directa los datos y un sistema los almacena en una base de datos. Estos datos se obtiene de la fuente primaria, no a partir de otros datos.

En el segundo caso los datos se producen a partir de otros datos como se muestra en la figura 3.2. En este escenario hay tres bases de datos denominadas BD1, BD2, BD3 y un usuario productor que selecciona datos de ellas y los procesa para crear una nueva base de datos, que almacena los datos posiblemente procesados y creados a partir de los almacenados en BD1, BD2 y BD3.

Es probable que para que los datos de la nueva base de datos puedan ser accedidos por el usuario se requiera de un sistema que los haga disponibles. Sin embargo, en la figura 3.2 se hace énfasis en la creación no en los detalles de la presentación.

3.1.1. Producción de datos de forma estática y dinámica

Cuando los datos se toman de la fuente una sola vez y a partir de ese momento la fuente deja de ser útil y no se vuelve a ocupar, se dirá que la producción de datos es de forma *estática*. Por el contrario, si la fuente de datos es utilizada de forma constante durante cierto período de tiempo se le llamará producción de datos de forma *dinámica* (continua).

Para comprender mejor la producción estática consideremos el caso de la figura 3.2 e imaginemos un productor de datos que está interesado en entregar una sola vez a un usuario un conjunto de datos. El usuario productor tomará los datos de la BD1, BD2 y BD3 y los procesará para crear una nueva base de datos. Cuando el productor tenga la nueva base de datos creadas no tendrá más relación con BD1, BD2 y BD3, por lo que si estas posteriormente son borradas o cambian su estructura no es relevante para el productor.

Ahora analicemos la producción dinámica de datos. Si consideramos nuevamente la figura

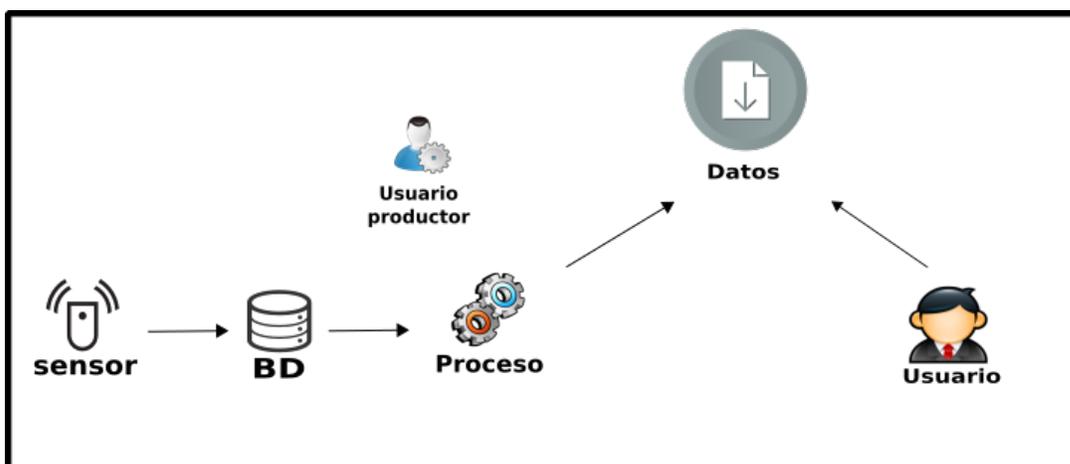


Figura 3.1: Escenario de captura de datos a través de un sensor.
Fuente: Creación propia

3.2 e imaginamos que el productor de datos está interesado en entregar de forma periódica a un usuario un conjunto de datos, este tomará los datos de la BD1, BD2 y BD3, luego de procesarlos creará una nueva base de datos, y a diferencia del caso anterior el productor de datos estará pendiente de la disponibilidad, de la coherencia en el tiempo, de la estandarización, de la sostenibilidad, de lo actualizadas que estén las BD1, BD2 y BD3. Es probable que el productor esté interesado también en muchas otras características de la base de datos que no se mencionaron. También en la producción estática hay características que la base de datos debe tener que no se han mencionado y que se abordarán más adelante como parte de las dimensiones de los datos.

Un ejemplo de productor estático es el que quiere conocer cuántos alumnos aprobaron en todas las escuelas de una comuna para aplicar una medida estadística (moda, promedio, etc.), comparar las comunas y entregar datos al respecto. En este caso, el productor lo hace solo una vez por lo que toma los datos y no le interesa si el siguiente año publicarán esos mismos datos o si hace dos años los publicaron. Por otro lado, para ilustrar la producción dinámica consideraremos el caso del productor de datos que toma datos del tiempo (temperatura, lluvia, humedad, etc.) de la Dirección meteorológica nacional cada hora y aplica una medida estadística para presentar la temperatura diaria en todo un año. En este caso, requiere los datos actualizados cada hora y requiere que la estructura en la que se entregan los datos de forma automática no cambie, entre otras cosas.

Los desarrolladores que utilizan datos abiertos crean productos y servicios. Entre sus productos se destacan tanto la creación de datos como de información. El procedimiento genérico para crear información es bastante similar al de crear datos a partir de datos, su principal diferencia es lógicamente el producto final. La figura 3.3 muestra el escenario de creación de información. La producción de información también se puede realizar en forma estática y dinámica, tal como se explicó en el escenario de los datos.

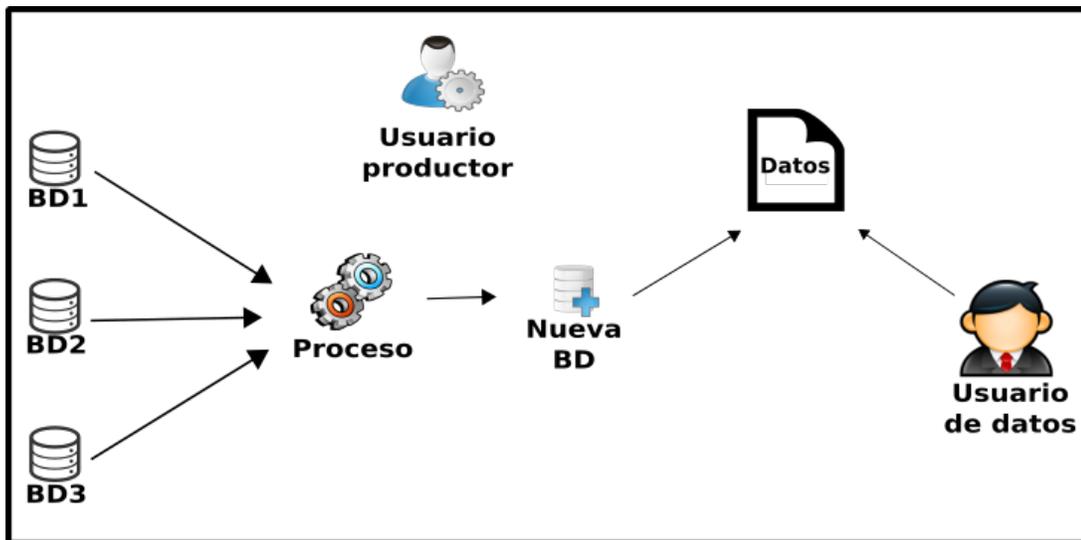


Figura 3.2: Escenario de creación de datos a partir de otros datos.
Fuente: Creación propia

3.1.2. Usuarios técnicos y no técnicos

En el contexto del uso de datos abiertos los usuarios productores pueden ser de diferentes tipos. Por ejemplo, están los gerentes comerciales, que poseen el capital y tienen la estrategia del negocio, pero no poseen los conocimientos técnicos. Estos contratan personal con los conocimientos para manipular los conjuntos de datos. Esto nos lleva a considerar dos posibles subtipos de productores, los *técnicos* y los *no técnicos*.

Por otro lado, los usuarios que consumen datos pueden ser usuarios finales. Por ejemplo, los ciudadanos que consumen los datos de forma secundaria, por medio de periódicos, mashups, revistas académicas, entre otros. Estos usuarios no tienen necesidad de tener conocimientos técnicos pues consumen los datos procesados. ¿Quiénes más consumen los datos abiertos? Pues los productores, que son al mismo productores y consumidores: ellos pueden requerir conocimiento técnico o no, tal como se analizó anteriormente. Esto nos permite reconocer que los consumidores también tienen dos posibles subtipos, los *técnicos* y los *no técnicos*, de acuerdo a la forma en que usen los datos.

3.1.3. Relaciones entre usos y usuarios

Anteriormente los usuarios de los datos se clasificaron en dos grandes grupos conocidos como usuarios productores y usuarios consumidores. Esta clasificación se encuentra ligada al uso que éstos le den a los datos; es decir, si solamente consumen los datos o si producen datos o información a partir de ellos. Una característica que influye de manera importante en este caso es que los datos sean primarios.

Considerando la importancia de la fuente de datos través del tiempo los usos se clasificaron en: usos de corto tiempo (o plazo) llamado acá *estático*. O de largo tiempo (o plazo) llamado

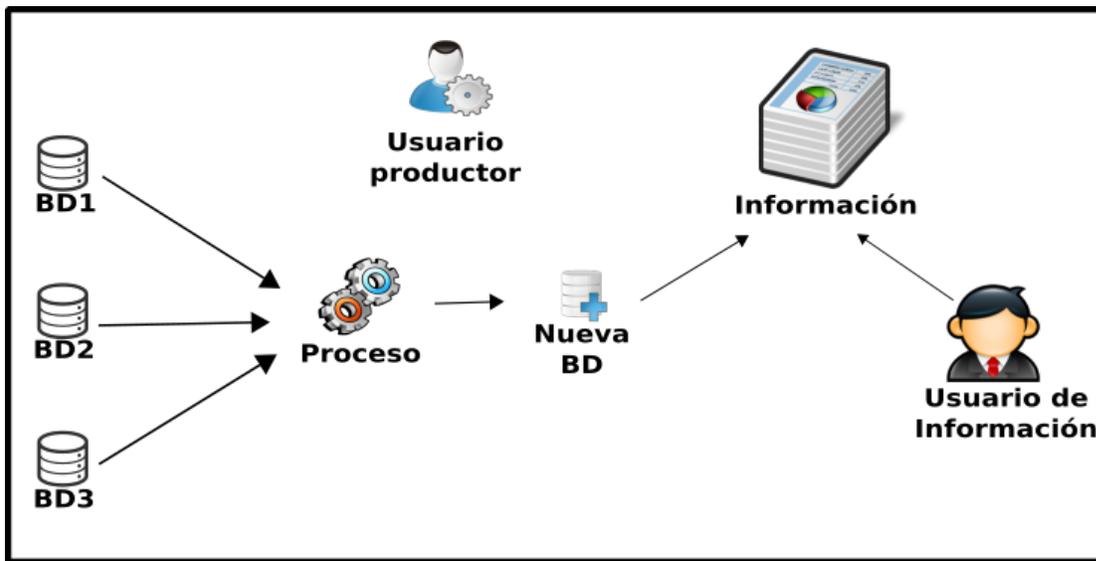


Figura 3.3: Escenario de productor de información.

Fuente: Creación propia

dinámico. En el primero, el usuario toma los datos crea un producto y no se vuelve a interesar en la fuente de datos. En el uso dinámico interesa tanto que la fuente produzca nuevos datos, como los mecanismos de consistencia, actualización, entre otras cosas. Estos usos se encuentran ligados a la producción o al consumo.

Desde otro punto de vista, los usuarios se pueden clasificar de acuerdo a las habilidades y destrezas que requieran para manipular o usar los datos. En este caso, conocemos los usuarios técnicos y no técnicos. De forma sencilla, el usuario técnico es aquel que para utilizar los datos y obtener valor de ellos utiliza conocimientos especializados en algún área, por ejemplo, estadística; computación; medicina; entre otros. Sin embargo, para este trabajo solo se considerarán las habilidades técnicas informáticas. El usuario no técnico es el que no utiliza habilidades especiales para el uso de los datos. (normalmente llamados “hombre de la calle”, “sra. Juanita”, etc.)

Para comprender mejor esta relación de los usos y los usuarios la figura 3.4 muestra en forma tridimensional las diferentes dimensiones de los usuarios y usos.

3.2. Dimensiones de datos

La definición de dimensión brindada por la Real Academia Española que más se relaciona con la forma en que se habla de dimensión en este contexto es “Aspecto o faceta de algo”. En este caso se refiere a qué aspectos, cualidades o características de los datos se van a tomar en cuenta para evaluar la apertura, capacidad de reutilización y accesibilidad. Los estudios acerca de los datos se han realizado desde diferentes perspectivas: desde la estadística [50], [30], desde las ciencias de la computación como se observó en el capítulo 2 puede estar relacionado con los procesos de visualización de datos [20], y con las bases de datos, que es de donde más se ha concentrado el trabajo y cuyas publicaciones se mencionan en el desarrollo

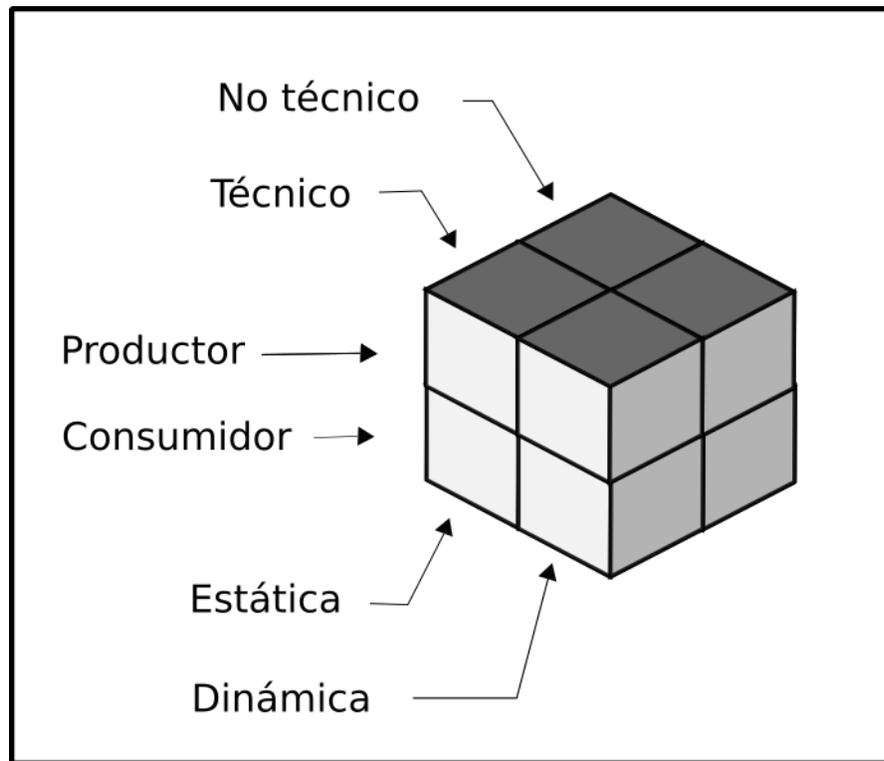


Figura 3.4: Resumen de las principales dimensiones de los usos y usuarios de datos abiertos.
Fuente: Creación propia

de esta sección.

Wang y Strong en 1996 presentan un trabajo [100] a partir del cual mucha de la literatura actual se basa. Ellos realizaron dos encuestas, la primera para recopilar las principales atributos de calidad de los datos. La segunda para evaluar la importancia de los atributos de calidad recopilados en la primera. Estas encuestas fueron aplicadas a consumidores de datos que utilizaban los datos para tomar decisiones en diferentes contextos. Con los resultados de la segunda encuesta y tras algunos procesos de refinamiento definen cuatro categorías que permiten clasificar los atributos de calidad que estudiaron, estas categorías son: *Intrínseco*, *Contextual*, *Representacional* y *Accesible*. De estas cuatro surgen quince dimensiones que se pueden observar en la figura 3.5.

La categoría *intrínseca* agrupa aquellas dimensiones que se consideran propias de los datos y se subdivide en: reputación, creíble, objetividad y exactitud.

La categoría *contextual* contiene las dimensiones que permiten evaluar los datos con respecto a la uso que se le dará y se subdivide en: valor agregado, relevancia, puntualidad/prontitud (Timeliness), completitud, apropiada cantidad de datos.

La categoría *representacional* es la categoría que reúne las dimensiones relacionadas con el formato y con el significado de los datos y se subdivide en: interpretable, fácil de entender, consistencia representacional, concisa.

La categoría *accesible* enfatiza la importancia de los sistemas donde residen los datos y se

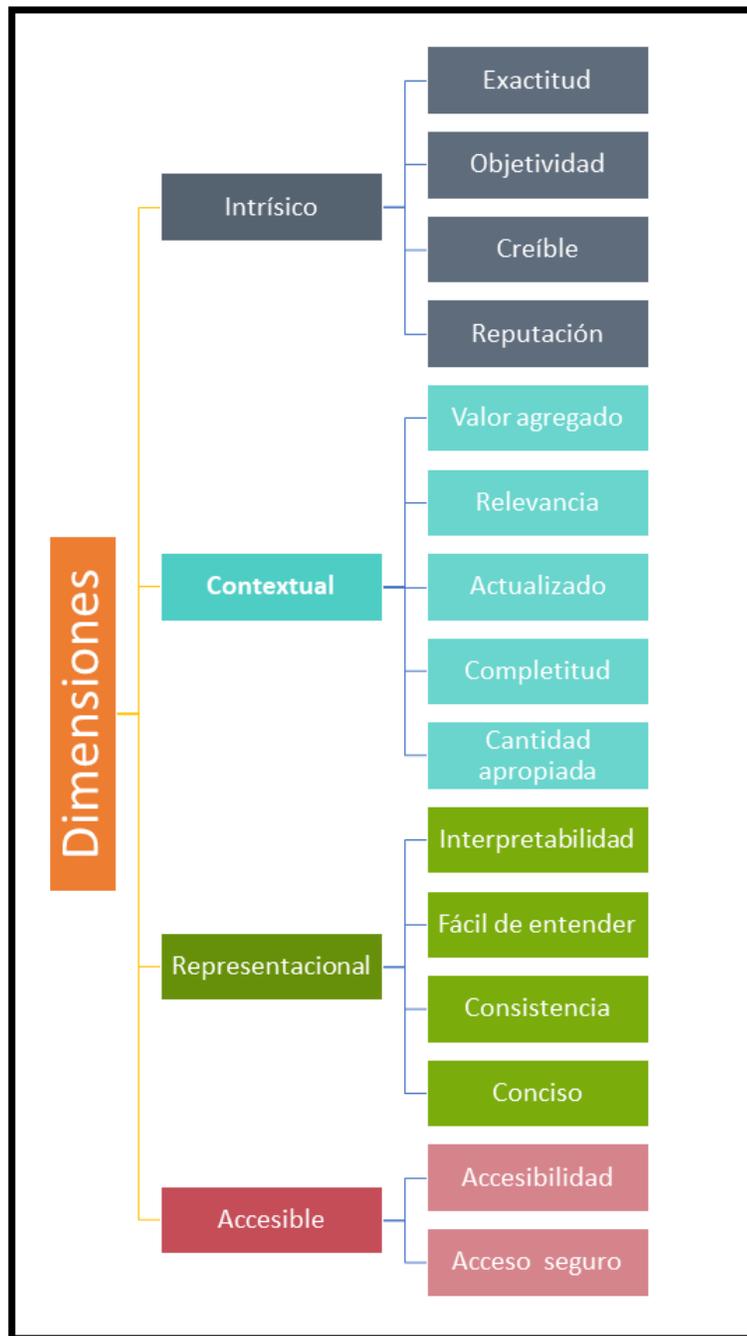


Figura 3.5: Dimensiones propuestas por Wang y Strong, 1996
 Fuente: Adaptado de Beyond accuracy: What data quality means to data consumer [100]

subdivide en: accesibilidad y seguridad de acceso.

Previo al muy citado trabajo realizado en 1996 [100], Richard Wang y Lisa Guarascio propusieron en un artículo [101] publicado 1991 la clasificación de la figura 3.6. Este trabajo también consistió en dos encuestas, la primera enfocada en determinar las dimensiones de calidad de los datos, la segunda encuesta varió levemente de la primera y permitió clasificar las dimensiones en: extremadamente importante, importante y no importante.

Value Added (2) Cost Effectiveness (14)			
Ease of Understanding(6)	Believability (1)	Relevancy (3)	Timeliness (9)
Interpretability (5) Representational Consistency (13) Conciseness (17)	Accuracy (4) Objectivity (8) Completeness (10) Traceability (11) Reputation (12) Variety of Data & Data Sources (16) Access Security (18)	Appropriate Amount of Data (19)	Accessibility (7) Ease of Operation (15) Flexibility (20)

Figura 3.6: Dimensiones propuestas por Wang y Guarascio, 1991.

Fuente: Dimensions of Data Quality: Toward Quality Data by Design [101]

Por otro lado, Thomas Redman y colegas en 1994 presentaron su trabajo “The notion of data and its quality dimensions” [40] en el que agruparon las dimensiones de valores en cuatro categorías: *accuracy*, *currentness*, *completeness* y *consistency*. Las dimensiones de modelo y representación de datos no las presentaron en ese trabajo, pero indicaron que ya habían abordado el tema en el capítulo tres de su libro “Data quality: Management and Technology” [80]. En 1995 Redman y Levitin [53] estudian seis aspectos de los datos: content, scope, level of detail, composition, consistency, and reaction to change, y los problemas relacionados con ellos.

Thomas Redman y sus colegas clasifican las dimensiones de datos de acuerdo al enfoque de estudio de cada trabajo. Por ejemplo, en el párrafo anterior se citaron las clasificaciones propuestas en los años 1994 [40] y 1995 [53], en la figura 3.8 se aprecian las clasificaciones propuestas en los trabajos “Data Quality for the Information Age” [82] publicado en 1997 y “Data Quality: The Field Guide” [81] publicado en el año 2001. En la figura 3.7 se puede observar la clasificación propuesta en el libro “Handbook of data quality: research and practice” [84] publicado en el año 2013.

Wang centra sus trabajos en obtener las propiedades más importantes de los datos a partir de los intereses de los usuarios. El trabajo desarrollado en 1996 se compone de dos encuestas realizadas a los consumidores de datos. La primera encuesta recopila las dimensiones y en la segunda evalúa la importancia de las dimensiones recopiladas, y con ello realiza un análisis exploratorio que concluyó con la clasificación que propuso. Por su parte, Redman aborda el tema de los datos modelando el mundo real a través de entidades, atributos y valores. Su enfoque está más orientado a las estructuras de los datos y al modelado de datos. Con esta perspectiva plantea los problemas del valor de los datos y extrae las dimensiones. En trabajos posteriores a [80] y [40] considera tres categorías: modelo de datos, valores de datos, representación de datos.

La literatura presenta una pluralidad de obras sobre atributos de datos [100], [55], [13], [99], [69], [34], entre otras. Los puntos de vista son variados, o sea que hay opiniones desde el punto de vista estadístico, de sistemas de información, de bases de de datos, entre otros. Las definiciones de las dimensiones no siempre coinciden entre los diferentes autores, pero lo que sí es claro es que cada definición está de acuerdo a su contexto, lo que alienta a hacer un

Conceptual view/associated metadata	Data values
Appropriate use	Accuracy
Areas covered	Completeness
Attribute granularity	Consistency
Clear definition	Timeliness
Comprehensiveness	
Essentialness	<i>Presentation quality</i>
Flexibility	Appropriateness
Homogeneity	Ease of interpretation
Identifiability	Formats
Naturalness	Format precision
Obtainability	Flexibility
Precision of domains	Handling of null values
Relevancy	Language
Robustness	Portability
Semantic consistency	Representation Consistency
Sources	Use of storage
Structural consistency	

Figura 3.7: Dimensiones propuestas por Thomas C. Redman.

Fuente: Handbook of data quality: research and practice [84]

análisis sobre las dimensiones en el contexto de datos abiertos. Las figuras 3.8 y 3.9 muestran un resumen de las propiedades de los que algunos autores consideran importantes, los autores han propuesto diferentes clasificaciones de acuerdo al enfoque de su estudio.

El grupo de trabajo del proyecto UNECE/HLG de la Oficina Nacional de Estadística de Canadá entregó un informe técnico [34] sobre la aptitud de uso de los datos, el cual ofrece la siguiente clasificación de variables:

- **Fuente:** entorno institucional/empresarial, privacidad y seguridad.
- **Metadatos:** complejidad, claridad de accesibilidad, integridad, usabilidad, factores de tiempo, coherencia (vinculación y consistencia), validez.
- **Datos:** Exactitud, coherencia (consistencia y vinculación), validez, factores de tiempo.

Research work	Structure	Terms used
Beyond accuracy: What data quality means to data consumers [15]	4 categories, 15 dimensions	Intrinsic: accuracy, objectivity, believability, reputation Contextual: value-added, relevancy, timeliness, completeness, appropriate amount of data Representational: interpretability, ease of understanding, representational consistency, concise representation Accessible: accessibility, access security
Anchoring data quality dimensions in ontological foundations [31]	2 views, 2 categories, 26 dimensions	Internal view Data-related: accuracy, reliability, timeliness, completeness, currency, consistency, precision System-related: reliability External view Data-related: timeliness, relevance, content, importance, sufficiency, usability, usefulness, clarity, conciseness, freedom from bias, informativeness, level of detail, quantitiveness, scope, interpretability, understandability System-related: timeliness, flexibility, format, efficiency
DoD Guidelines... [26]	6 characteristics	Accuracy, completeness, consistency, timeliness, uniqueness, validity
Data Quality Handbook... [42]	5 dimensions	Correctness, completeness, consistency, currency, accessibility
Data Quality for the Information Age [32]	3 categories, 27 dimensions (15 characteristics)	Data model: content (data relevance, obtainability of values, clarity of definition), scope (comprehensiveness, essentialness), level of detail (attribute granularity, domain precision), composition (naturalness, identifiability, homogeneity, minimal unnecessary redundancy), model consistency (structural, semantic consistency), reaction to change (flexibility, robustness) Data values: accuracy, completeness, currency, and consistency Data representation: appropriateness, interpretability, portability, format precision, format flexibility, ability to represent null values, efficient usage of recording media, representation consistency
Improving data warehouse and business information quality [33]	2 categories, 15 dimensions	Inherent: definition conformance, value completeness, validity/business rule conformance, accuracy to surrogate source, accuracy to reality, precision, non-duplication, equivalence/concurrency of redundant or distributed data Pragmatic: accessibility, timeliness, contextual clarity, derivation integrity, usability, rightness or fact completeness
Assessment methods for information quality criteria [34]	3 classes, 22 criteria	Subject: believability, concise representation, interpretability, relevancy, reputation, understandability, value-added Object: completeness, customer support, documentation, objectivity, price, reliability, security, timeliness, verifiability Process: accuracy, amount of data, availability, consistent representation, latency, response time
Data Quality: The Guide Field [41]	9 categories, 51 dimensions	Accessibility/Delivery: availability, protocol, security Quality of content: attribute granularity, comprehensiveness, essentialness, flexibility, appropriate use, areas covered, homogeneity, naturalness, obtainability, precision of domains, robustness, semantic consistency, structural consistency, simplicity, clear definition, identifiability, source, relevancy Quality of values: accuracy, completeness, timeliness, consistency Presentation quality: appropriateness, format precision, use of storage Flexibility: portability, representation consistency, null values, formats, language, ease of interpretation Improvement: feedback, measurement, track record Privacy: consumer privacy, privacy of others, security Commitment: warning, help, special requests, commitment Architecture: library/documentation, logical structure, physical structure, naming, rules, redundancy, unit cost
Enterprise Knowledge Management: The Data Quality Approach [36]	5 categories, 31 characteristics	Data models: clarity, comprehensiveness, flexibility, robustness, essentialness attribute granularity, domain precision, homogeneity, naturalness, identifiability, obtainability, relevance, simplicity, semantic/structural consistency Data values: accuracy, null values, completeness, consistency, currency/timeliness Information domains: enterprise agreement of usage, stewardship, ubiquity Data presentation: appropriateness, correct interpretation, flexibility, format precision, portability, representation consistency, representation of null values, use of storage Information policy: accessibility, metadata, privacy, redundancy, security, unit cost
Data quality assessment [12]	16 dimensions	Accessibility, appropriate amount of data, believability, completeness, concise/consistent representation, ease of manipulation, free-of-error, interpretability, objectivity, relevancy, reputation, security, timeliness, understandability, value-added
A New Method for Database Data Quality Evaluation at the Canadian Institute... [37]	5 categories 24 characteristics	Accuracy, timeliness, comparability, usability, relevance Over-coverage, under-coverage, simple response variance, reliability, correlated response variance, collection and capture, unit and item non-response, edit and imputation, processing, estimation, timeliness, comprehensiveness, integration, standardization, equivalency, linkage-ability, product comparability, historical comparability, accessibility, documentation, interpretability, adaptability, value

Figura 3.8: Categorización de propiedades de datos en la literatura, parte 1.

Fuente: A survey on data quality: classifying poor data [52]

Research work	Structure	Terms used	
AIMQ: a methodology for information quality assessment [38]	4 quadrants (2 views plus 2 views)	Product quality	Conforms to specifications (Sound): free-of-error, concise representation, completeness, consistent representation Meets or exceeds consumer expectations (Useful): Appropriate amount, relevancy, understandability, interpretability, objectivity
		Service Quality	Conforms to specifications (Dependable): Timeliness, security Meets or exceeds consumer expectations (Usable): Believability, accessibility, ease of operation, reputation
		Accessibility	
A conceptual framework and belief-function approach to assessing overall information quality [39]	4 attributes, 8 subattributes		Interpretability: Intelligible, meaningful Relevance: user-specified criteria, timely (considers age and volatility) Credibility: accuracy, completeness, consistency, and non-fictitiousness
			Intrinsic: Accuracy/validity, cohesiveness, complexity, semantic consistency, structural consistency, currency, informativeness / redundancy, naturalness, precision / completeness
			Relational or contextual: Accuracy, accessibility, complexity, naturalness, informativeness / redundancy, relevance (aboutness), precision / completeness, security, semantic consistency, structural consistency, verifiability, volatility Reputation: authority
A Framework for Information Quality Assessment [40]	3 categories, 22 dimensions		Inherent: accuracy, completeness, consistency, credibility, currentness; accessibility, compliance, confidentiality, efficiency, precision, traceability, understandability System dependent: availability, portability, recoverability Inherent and system dependent: accessibility, compliance, confidentiality, efficiency, precision, traceability, understandability
			Intrinsic: accuracy, credibility (objectivity, reputation), traceability, currentness, expiration, completeness, consistency, accessibility (interactive, operation ease, cust. support), compliance, confidentiality, efficiency, precision, understandability Operational: availability, accessibility, verifiability, confidentiality, portability, recoverability
			Contextual: validity (reliability, scope), value-added (applicability, flexibility, novelty), relevancy (novelty, timeliness), specialization, usefulness, efficiency, effectiveness, traceability, compliance, precision Representational: concise representation, consistent representation, understandability, (interpretability, amount of data, documentation, organization), attractiveness, readability
ISO/IEC 25012 [8]	2 points of view and 1 mixed view, 15 characteristics		Intrinsic: Accuracy, Lineage, Semantic, Structure. Contextual: completeness, consistency, currency, timeliness, reasonableness and identifiability Qualitative dimensions
			Process pragmatics: ease of operation, ease of navigation, interactivity, flexibility of representation, suitability of representation, access security, accessibility, latency, response time, availability Syntactics: consistency, semantic consistency, structural consistency, conformability, naturalness, integrity User pragmatics: believability, verifiability, amount empirical evidence, reliability, reputation, user-conformability, enjoyability Semantics: informativeness, conciseness, accuracy, objectivity, currency, completeness, cohesiveness, degree of context, maintainability, unambiguous Information pragmatics: understandability, interpretability, usability, efficiency, value add, complexity, relevancy, timeliness, volatility
			Accuracy: accuracy, unbiased, believability, traceability Accessibility: accessibility, believability, appropriate amount of information Usefulness: interpretability, understandability, ease of manipulation, consistent representation, value added Relevance: relevant, concise representation, up-to-date, reputation, value added Security: security, traceability
The Practitioner's Guide to Data Quality Improvement [17]	3 categories, 10 Characteristics		Completeness, accuracy, correctness, consistency, and timeliness
			Source: institutional/business environment, privacy and security Metadata: complexity, accessibility clarity, completeness, usability, time factors, coherence(linkability and consistency), validity Data: accuracy (selectivity), coherence (consistency and linkability), validity, time factors
			Contextual consistency Velocity: consistency, credibility, confidentiality Volume: completeness, credibility Variety: accuracy, consistency, understandability
Information Quality Dimensions for the Social Web [45]	5 categories, 42 criteria		Temporal consistency Velocity: consistency, credibility, currentness, availability Volume: availability Variety: consistency, currentness, compliance
			Operational consistency Velocity: completeness, accessibility, efficiency, traceability, availability, recoverability Volume: completeness, accessibility, efficiency, availability, recoverability Variety: accuracy, accessibility, compliance, efficiency, precision, traceability, availability,
Prioritization of data quality dimensions and skills requirements in genome annotation work [46]	5 constructs, 19 dimensions		
Towards an ontology... [47]	5 dimensions		
A Suggested Framework for the Quality of Big Data [48]	3 hyperdimensions, 11 dimensions, 3		
A Data Quality in Use Model for Big Data [49]	3 by 3 views, 15 characteristics		

Figura 3.9: Categorización de propiedades de datos en la literatura, parte 2.

Fuente: A survey on data quality: classifying poor data [52]



Figura 3.10: Esta matriz relaciona las expectativas de usuarios propuestas por Redman y las funcionalidades de los portales Web propuestas por Collins, es parte del trabajo presentado por Caro y otros, en 2008
 Fuente: Caro y colegas, 2008 [15]

3.3. Dimensiones de calidad de datos de un portal Web

Caro y otros, propusieron en el año 2008 [15] un conjunto de atributos para la evaluación de la calidad de datos de un portal Web, denominado Portal Data Quality Model (en adelante PDQM). Ellos indican que las dimensiones que proponen evalúan la calidad de datos de un portal Web desde el punto de vista del consumidor. También indicaron que para identificar dichos atributos usaron tres elementos básicos: primero, un conjunto de atributos de calidad de datos Web identificados en la literatura. Segundo, las expectativas de datos que tienen los consumidores de datos en Internet, descritas por Redman [81]. Tercero, las funciones que un portal puede ofrecer a sus usuarios, según lo descrito por Collins [24].

Su idea es reflejar el hecho de que un usuario evalúa la características de calidad de los datos cuando usa las funciones y servicios que el portal Web ofrece. Para ello construyeron la matriz de la figura 3.10, donde relacionan las expectativas de calidad de datos que tienen los consumidores de datos en Internet propuestas por Redman [81] con las funciones que un portal puede ofrecer a sus usuarios que propuso Collins [24], en dicha figura marcan las intersecciones que ellos consideran válidas o que tienen sentido. Seguidamente, clasifican las dimensiones encontradas en la literatura dentro de las matriz. La figura 3.11 muestra el resumen que ellos entregaron de la literatura revisada.

Posteriormente, ellos validan los atributos de calidad de datos por medio de una encuesta a consumidores de datos, buscando obtener el nivel importancia de cada atributo y eliminar los menos relevantes. La validación les entregó 33 atributos de calidad de datos que estructuraron como se muestra en la figura 3.12.

En 2009 Moraga y sus colegas [65] presentaron un modelo de calidad de datos para portales Web, denominado SQuaRE-Aligned Data Quality Model for Web Portals (en adelante SPDQM) de este se basó en el PDQM (Portal Data Quality Model) y el estándar de calidad de datos ISO/IEC 25012. El PDQM es el trabajo propuesto por Caro y otros [15] que se describió anteriormente.

El modelo de calidad de producto de datos definido en el estándar ISO/IEC 25012 está

Autor	Domain	No. of DQ attributes obtained from the model/framework
Katerattanakul and Siau (1999)	Personal web sites	6 DQ attributes
Katerattanakul and Siau (2001)	e-Commerce	
Naumann and Rolker (2000)	Data integration	22 DQ attributes
Pernici and Scannapieco (2002)	Web information systems (data evolution)	4 DQ attributes
Fugini et al. (2002)	e-Service cooperative	8 DQ attributes
Graefe (2003)	Decision making	8 DQ attributes
Eppler et al. (2003)	Web sites	16 DQ attributes
Gertz et al. (2004)	DQ on the web	5 DQ attributes
Moustakis et al. (2004)	Web sites	4 DQ attributes
Melkas (2004)	Organizational networks	20 DQ attributes
Bouzeghoub and Peralta (2004)	Data integration	2 DQ attributes
Yang et al. (2004)	Web information portals	5 DQ attributes

Figura 3.11: Esta figura resume las fuentes de la revisión de la literatura realizada por Caro y otros, en 2008

Fuente: Caro y colegas, 2008 [15]

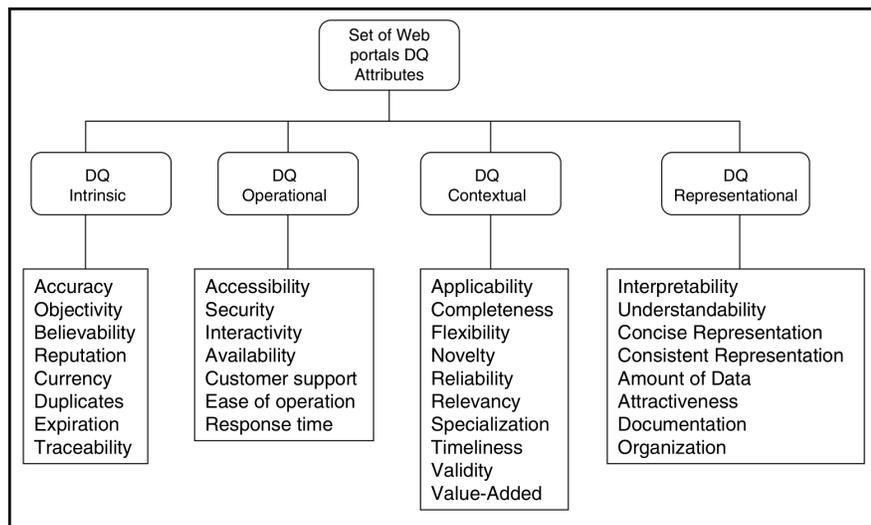


Figura 3.12: Esta figura muestra la estructura de las dimensiones de calidad de datos en portales Web planteada por Caro y otros, en 2008

Fuente: Caro y colegas, 2008 [15]

compuesto de 15 características, estas características se clasifican en dos categorías. La primer categoría es *inherentes a los datos*, y se refiere a características que son propias de los datos. En particular a: Valores de dominios de los datos y restricciones, relaciones entre los valores y metadatos. La segunda categoría es *dependientes del sistema*, se encuentra ligada al hardware y software que componen el sistema donde están alojados los datos. La figura 3.13 muestra la clasificación de las características en sus categorías.

En la primera etapa de SPDQM ellos realizaron la recolección de un conjuntos de características de calidad de datos. Para ello complementaron el trabajo realizado por los mismo autores [66], el PDQM y el estándar ISO/IEC 25012. El trabajo realizado por los autores consistió en una revisión de la literatura sobre características de calidad de datos de portales Web en el período comprendido entre 01/01/2006 y 31/12/2008. Ellos indican que encontra-

Characteristics	Data quality	
	Inherent	System dependent
Accuracy	X	
Completeness	X	
Consistency	X	
Credibility	X	
Currentness	X	
Accessibility	X	X
Compliance	X	X
Confidentiality	X	X
Efficiency	X	X
Precision	X	X
Traceability	X	X
Understandability	X	X
Availability		X
Portability		X
Recoverability		X

Figura 3.13: Esta figura muestra la clasificación de características de calidad de datos propuesta por la el estándar ISO/IEC 25012
Fuente: ISO/IEC 25012. Citada por Moraga y colegas, 2009 [65]

ron en la literatura algunas características con el mismo nombre que se referían a conceptos diferentes. Además algunos conceptos llamados iguales pero con significados diferentes. Entonces proponen una solución a estos conflictos adecuando las características al contexto de calidad de datos en portales Web. Posteriormente, organizan los datos como se detalla en la figura 3.14, De forma similar al estándar ISO/IEC 25012 clasifica las características en dos categorías: inherentes y dependientes del sistema, pero internamente presenta una subdivisión.

3.4. Dimensiones de datos abiertos

En la literatura revisada el primer trabajo que propone evaluar el grado de reutilización de los datos publicados en los portales de datos abiertos es Meloda, propuesto por Abella y otros [1] en el año 2014. Su trabajo tiene como base los 8 principios propuesto por la página de Datos Abiertos de Gobierno [74] y las tres leyes de datos abiertos propuestas por Eaves [36]. Eaves defina las siguientes tres leyes: ser accesible por Internet, ser legible por máquinas y que el marco legal permita usos distintos de los originales. Meloda analiza cuatro dimensiones (cito textualmente):

1. Estándares técnicos o estructura técnica en que se ofrece el dato
2. Acceso a la información o mecanismo por el cual se hace posible la descarga o conexión con la información.

Point of view	Category	Characteristic	Subcharacteristic
Inherent	Intrinsic: this denotes that data have quality in their own right	Accuracy:	Duplicates
		Credibility:	Objectivity
			Reputation
			Traceability
		Currentness	
		Expiration	
		Completeness	
		Consistency	
		Accessibility	
		Compliance	
		Confidentiality	
		Efficiency	
		Precision	
		Understandability	
System Dependent	Operational: this emphasizes the importance of the role of systems; that is, the system must be accessible but secure	Availability:	Response Time
		Accessibility:	Interactive
			Ease of operation
			Customer Support
		Verifiability	
		Confidentiality	
		Portability	
	Recoverability		
	Contextual: this highlights the requirement which states that data quality must be considered within the context of the task in hand	Validity:	Reliability
		Value-added:	Scope
			Applicability
			Flexibility
		Relevancy:	Novelty
		Specialization	Novelty
			Timeliness
		Usefulness	
		Traceability	
	Compliance		
	Precision		
	Representational: this denotes that the system must present data in such a way that they are interpretable, easy to understand, and concisely and consistently represented	Concise Representation	
		Consistent Representation	
		Understandability:	Interpretability
Amount of data			
Documentation			
Attractiveness:		Organization	
		Organization	
Readability			
Efficiency			
Effectiveness			

Figura 3.14: Esta figura muestra la estructura de las dimensiones de calidad de datos planteada en el SPDQM por Moraga y colegas, en 2009
Fuente: Moraga y colegas, 2009 [65]

3. Marco legal: licencia que se asigna al conjunto de datos.
4. Modelos de datos: modelo utilizado para publicar la información y la cantidad de veces que es utilizado en otras fuentes de datos

Ellos proponen una evaluación cuantitativa, dan un valor a cada dimensión y luego calculan el puntaje final como la raíz cuarta del producto de las puntuaciones obtenidas en cada dimensión. Los niveles que evalúan en cada dimensión se resumen en la figura 3.15.

Vetrò y colegas en el 2016 [97] proponen un conjunto de características para evaluar la calidad de los datos abiertos. Ellos tomaron como principal apoyo para formar su marco teórico el trabajo realizado por Moraga y colegas, 2009 [65], y un survey realizado en el 2013, entre los participantes de una Hackatons. Utilizaron las respuestas de 15 desarrolladores a las siguientes dos preguntas: “Which problems did you find working with open data?” y “Which aspects of data quality would you like to improve?”. Con estas respuestas construyeron una lista de problemas (ver figura 3.16) y los relacionaron con las características del modelo SPDQM. De esta forma obtuvieron subconjunto de características vinculadas a los problemas de calidad de datos abiertos. La figura 3.17 muestra las características, métrica y definiciones propuestas por Vetrò y colegas.

Estándares técnicos	Acceso	Legal	Modelo de datos
1. Estándar privativo Ej.: .xls, .shp, .doc	1. Sin acceso Ej.: mail no automático o acceso en persona	1. <i>Copyright</i> Ej.: <i>copyright</i>	1. Sin modelo publicado Ej.: tabla de datos sin descripción de los campos
2. Estándar abierto Ej.: .csv, .ods, .wms	2. Acceso vía web con registro Ej.: formulario manual	2. Uso privado Ej.: <i>copyright</i> permitiendo uso personal	2. Modelo con campos de datos Ej.: tabla de datos con descripción de los campos
3. Estándar abierto con metadatos Ej.: .rdf, .rss, .json	3. Acceso directo vía web Ej.: url único	3. Uso no comercial Ej.: CC BY-NC 4.0	3. Modelo con especificaciones de campos Ej.: vocabularios disponibles
	4. Acceso vía web con parámetros Ej.: url con parámetros	4. Uso comercial Ej.: CC BY-SA 4.0	4. Modelo externo normalizado Ej.: vocabularios disponibles aceptados por organización de normalización
	5. Acceso completo (API) Ej.: punto de acceso <i>Sparql</i>	5. Uso no limitado con autoría Ej.: CC BY 4.0	5. Modelo externo y generalizado Ej.: vocabularios disponibles aceptados por organización de normalización y reconocidos

Figura 3.15: Esta figura muestra las dimensiones y subdimensiones (niveles) propuestas por Abella, en 2014
Fuente: Abella, 2014 [1]

Oviedo [77] presenta un modelo de madurez para portales de datos abiertos, el cual consiste en un modelo cualitativo que toma como base el esquema de 5 estrellas de Berners-Lee [10] y lo amplía considerando el concepto de apertura en los datos. Considera tanto la calidad de los datos como la calidad de los portales que albergan los datos. El modelo considera los siguientes cinco criterios.

1. Disponibilidad de los datos publicados
2. Capacidad de reutilización de los datos publicados
3. Relevancia de los datos publicados
4. Granularidad de los datos publicados
5. Visualización de los datos publicados

De los trabajos estudiados en la revisión de la literatura ninguno considera las dimensiones de usuarios ni de usos. En el siguiente capítulo se proponen un conjunto de dimensiones a partir de la literatura estudiada y considerando las dimensiones de uso y usuarios.

3.5. Dimensiones de datos abiertos según usos y usuarios

Ya se analizaron varios conjuntos y clasificaciones de dimensiones de los datos y de datos abiertos, algunos de los trabajos centraban su atención en datos de sistemas de información o bases de datos. Otros proponen algunas dimensiones para datos abiertos, pero ninguno considera las metadimensiones de usuarios y usos. Por otro lado, los trabajos presentados por Piet Daas y sus colegas [30] y por Karr y compañía [50] se refieren al uso de los conjuntos de datos. Fierro [38], Mora [64] y Lanza [51] como ya se mencionó anteriormente proponen dimensiones para la creación de índices de información de transparencia. *Ninguno realiza un acercamiento a los datos abiertos considerando las metadimensiones de usuarios y usos.* En

Problem found in survey	Related quality characteristic	Intrinsic	System-dependent
Incomplete data	Completeness	X	
Format not compliant to well known standard	Compliance		
Lack of data source traceability	Traceability	X	X
Incongruent data	Consistency	X	
Out-of-date data	Expiration, currentness	X	
Lack of Metadata	Compliance, understandability		
Errors	Accuracy	X	
High time to understand data	Understandability	X	X

Figura 3.16: Esta figura muestra los principales problemas detectados en la encuesta realizada en los Hackatons, según el trabajo realizado por Vetrò y colegas, en 2016
Fuente: Vetrò y colegas, 2016 [97]

este trabajo, tomando esa experiencia y observando los usos, los usuarios y las características particulares de los datos abiertos se ha construido una clasificación, un esquema que permite evaluar la apertura de los datos, la accesibilidad y el nivel de reutilización de los datos, así como conocer a qué tipo de usos/usuarios están dirigidos los datos publicados.

Se debe considerar que los datos abiertos tienen en general características distintas a los de una base de datos relacional. Por ejemplo, la estricta disponibilidad web, la apertura legal, el tamaño, la capacidad de ser fácilmente portable (esto implica un tamaño adecuado para transportarse en la Web, y un formato abierto para almacenarse y transportarse), el nivel de desagregación, la necesidad de entregarse en una forma que promueva la interoperabilidad, entre otros. Esto además de muchas de las características de las bases de datos relacionales que también poseen los datos abiertos.

Para la elaboración del esquema propuesto se complementaron la dimensiones encontradas en la revisión de la literatura y las de la recopilación realizada por Laranjeiro y colegas [52] que se muestran en las figuras 3.8 y 3.9. La literatura revisada se encuentra resumida en la figura 3.18. En la parte superior de dicha figura (es decir por encima de la línea de tiempo) se encuentran los trabajos estudiados que se relacionan con calidad de datos y dimensiones de datos abiertos. En la parte inferior (por debajo de la línea de tiempo) se encuentran los trabajos relacionados con evaluación de la información transparencia, entre ellos algunos índices de evaluación de transparencia desarrollados en los países latinoamericanos estudiados.

Después de estudiar las dimensiones propuestas en la literatura se eliminaron las repetidas, con especial cuidado ya que existían algunas con diferente nombre y que se referían a lo mismo, por ejemplo, “comprensible” y “fácil de entender”. Por otro lado, habían otras dimensiones con el mismo nombre pero en diferentes contextos, por ejemplo, exactitud del valor de un atributo en un conjunto de datos, con respecto al valor en el mundo real y exactitud del valor con respecto a una fuente original (para el caso de una fuente replicada). Partimos del hecho que no existe un consenso en la literatura sobre una definición exacta de cada dimensión que la determine, las mismas son borrosas, esto es tienen límites difusos.

Characteristic	Metric	Level	Description
Traceability	Track of creation	Dataset	Indicates the presence or absence of metadata associated with the process of creation of a dataset.
	Track of updates	Dataset	Indicates the existence or absence of metadata associated with the updates done to a dataset.
Currentness	Percentage of current rows	Cell	Indicates the percentage of rows of a dataset that have current values, it means that they don't have any value that refers to a previous or a following period of time.
	Delay in publication	Dataset	Indicates the ratio between the delay in the publication (number of days passed between the moment in which the information is available and the publication of the dataset) and the period of time referred by the dataset (week, month, year).
Expiration	Delay after expiration	Dataset	Indicates the ratio between the delay in the publication of a dataset after the expiration of its previous version and the period of time referred by the dataset (week, month, year).
Completeness	Percentage of complete cells	Cell	Indicates the percentage of complete cells in a dataset. It means the cells that are not empty and have a meaningful value assigned (i.e. a value coherent with the domain of the column).
	Percentage of complete rows	Cell	Indicates the percentage of complete rows in a dataset. It means the rows that don't have any incomplete cell.
Compliance	Percentage of standardized columns	Cell	Indicates the percentage of standardized columns in a dataset. It just considers the columns that represent some kind of information that has standards associated with it (i.e. geographic information).
	eGMS Compliance	Dataset	Indicates the degree to which a dataset follows the e-GMS standard (as far as the basic elements are concerned, it essentially boils down to a specification of which Dublin Core metadata should be supplied)
	Five star Open Data	Dataset	Indicates the level of the 5 star Open Data model in which the dataset is and the advantage offered by this reason.
Understandability	Percentage of columns with metadata	Cell	Indicates the percentage of columns in a dataset that has associated descriptive metadata. This metadata is important because it allows to easily understanding the information of the data and the way it is represented.
	Percentage of columns in comprehensible format	Cell	Indicates the percentage of columns in a dataset that is represented in a format that can be easily understood by the users and it is also machine-readable.
Accuracy	Percentage of accurate cells	Cell	Indicates the percentage cells in a dataset that has correct values according to the domain and the type of information of the dataset.
	Accuracy in aggregation	Cell	Indicates the ratio between the error in aggregation and the scale of data representation. This metric only applies for the datasets that have aggregation columns or when there are two or more datasets referring to the same information but in a different granularity level.

Figura 3.17: Esta figura muestra las métricas propuestas por Vetrò y colegas, en 2016
Fuente: Vetrò y colegas, 2016 [97]

En un contexto donde los sistemas digitales y la tecnología de información y comunicaciones han venido desempeñando un papel protagónico, surge la necesidad de que dos o más sistemas independientes se puedan comunicar. Más que comunicar dos sistemas se busca intercambiar datos, integrar servicios de forma ágil y que los usuarios no enfrenten complicaciones. Para alcanzar esto es necesario ponerse de acuerdo en las estructuras que permitirán el intercambio. En particular, el uso de estándares abiertos, y el llegar a consensos en arquitecturas considerando el desarrollo en módulos que ofrezcan flexibilidad.

El párrafo anterior nos ayudó a acercarnos al problema por el cual es necesario la interoperabilidad, entendiendo ésta como la capacidad o habilidad que deben tener los sistemas o componentes para interactuar, compartir datos de forma electrónica/digital a través del tiempo, valiéndose de la definición de estándares y consensos establecidos. La Comisión Europea reconoció en el 2004 tres dimensiones de la interoperabilidad: técnica, semántica y organizativa.

La interoperabilidad es una características deseada por las organizaciones en la que las propiedades de los datos cumplen un papel fundamental. La interoperabilidad no se considera como una propiedad de los datos. Por otro lado, algunas propiedades de los datos facilitan la interoperabilidad. Por ejemplo, se busca que los datos tengan metadatos, que sean consistentes, coherentes y que estén estandarizados, para contribuir a que haya interoperabilidad. Otras características que ayudan en la búsqueda de la interoperabilidad son: la capacidad de relacionar las fuentes de datos e identificarlos a través de identificadores únicos de recursos. También que los datos sean portables, a través de descargas de archivos o por medio de una API. En el caso de los archivos descargables el volumen de datos del archivo es un atributo que se debe conocer y que debe ser apropiado para que este sea *accesible*. El hecho de que los datos se encuentren accesibles en la Web sin requerir ningún tipo de registro y sin discriminación es también valorarle en lo que respecta a la *accesibilidad*.

La primera dimensión considerada fue *completitud*, esta citada por la NSF [68], y por Opengovdata [74] (ambos se refieren al término como completo no completitud) como parte de los principios de los datos abiertos. Redman [40] dice que la completitud es “el grado en que una colección de datos tiene valores para todos los atributos de todas las entidades que se supone que tienen valores”. Larry English [37] se refiere a la completitud como “La característica de tener todos los valores requeridos para los campos de los datos”. Considerando las definiciones se puede concluir que dicha característica es importante en los conjuntos de datos abiertos. Por ejemplo, si se desea tener los datos de salud de todos los niños entre cinco y diez años, pero hay un alto porcentaje de registros que tienen nulo el campo de fecha de nacimiento, probablemente la estadística será incorrecta. En este sentido, la certeza de que los datos estén completos es un factor relevante para técnicos, como para los no técnicos.

La segunda dimensión es *procesable por máquina* y es mencionada por Opengovdata [74] como parte de las características que deben tener los datos abiertos. El esquema propuesto por Berners-Lee [10] ofrece dos estrellas si los datos abiertos son publicados en formato procesable por máquina. El Open Data White Paper [61] publicado por Reino Unido señala en su segundo punto “Los datos públicos se publicarán en forma reutilizable y legible por máquina”. Esta característica se encuentra relacionada con el formato que posee el archivo en el que se guardan los conjuntos de datos y con la estructura interna que estos poseen. Esta dimensión es fundamental para los usuarios técnicos, por tanto es la segunda dimensión que consideramos para los datos abiertos.

La tercera dimensión es *primario*, esta es citada por Wand y Wang [99] como nivel de detalle y por Redman [84] como atributo de granularidad. También es mencionada en Opengovdata [74] y de ella se dice lo siguiente: “Los datos se debe hacer disponibles de la forma más granular posible, no en forma agregada o modificada”. Según Loshin [56] la granularidad se refiere “al número de objetos que se utilizan para representar un concepto. Si un campo de datos puede tener varios valores, se dice que tiene una granularidad fina”. En este sentido, lo que se busca es tener un mayor nivel de detalle.

Respecto a las características primario o granular, que denota, por ejemplo, que los datos no están en porcentaje o en promedio, sino que están en la forma más granular o absoluta, por ejemplo, desagregado por sexo, por edad, por región o provincia. Estas características se encuentran tanto en los productores como los consumidores. Es deseable en los consumidores, pero vital para los productores, pues ofrece más posibilidades de crear. En este sentido, se puede plantear que la dimensión primario tiene una relación con la *capacidad del conjunto de datos para ser reutilizado*.

La cuarta dimensión considerada fue *estandarizado*, esta característica es mencionada por [9] y por Long y Seko [55]. Además es una característica importante para que sea posible interoperar, es decir, que haya interoperabilidad entre sistemas. En el contexto de datos Loshin [56] dice que un estándar “es un acuerdo entre las partes sobre las definiciones de términos comerciales comunes y la forma en que esos términos son nombrados y representados en los datos”. En este trabajo se aborda considerando que los datos abiertos deben considerar los estándares internacionales sobre monedas y medidas, entre otras cosas.

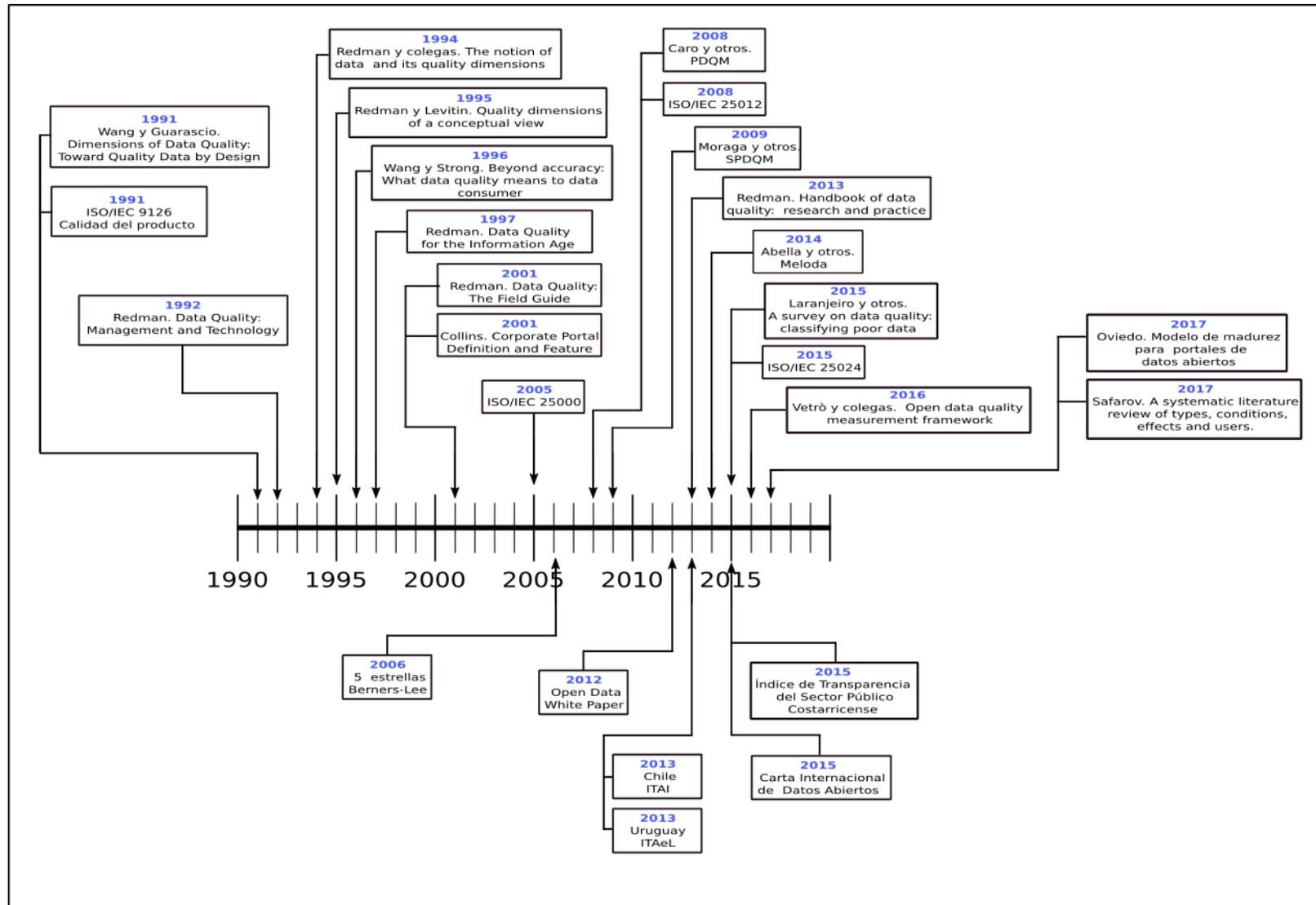


Figura 3.18: Esta imagen contiene una línea de tiempo con la literatura de dimensiones de: calidad de información de transparencia, calidad de datos y de datos abiertos

Fuente: Creación propia

La *portabilidad* fue la quinta dimensión que se incorporó al conjunto de dimensiones de datos abiertos propuestas, mencionada por Redman [82, 81] y por Loshin [56]. La norma ISO/IEC 25012 define portabilidad como el “grado en el que los datos tienen atributos que les permiten ser instalados, reemplazados o eliminados de un sistema a otro, preservando el nivel de calidad en un contexto de uso específico”. En el contexto de los datos abiertos esta dimensión evaluó la posibilidad de movilizar o exportar los datos.

La linkage-ability o habilidad para relacionarse se menciona en la literatura por Long y Seko [55]. Dufty y colegas [34] la define como “la facilidad con la que los datos pueden vincularse o fusionarse con otros conjuntos de datos relevantes”. El esquema propuesto por Berners-Lee [10] ofrece cinco estrellas si los datos abiertos son publicados con algunas características y además tiene la capacidad de relacionar con otros conjuntos de datos. Si se considera la transformación hacia la Web de datos que se está viviendo esta característica en los datos abiertos es relevante y necesaria. Por ello, la *capacidad de relacionarse* es la sexta dimensión incluida en este estudio.

La cantidad adecuada de datos es una característica importante en un archivo de datos y claramente en un conjunto de datos abiertos. Wang y Strong [100] dicen que “cantidad apropiada de datos es la medida en que la cantidad o el volumen de datos disponibles es apropiado”. Algunas características relacionadas a la dimensión *cantidad adecuada de datos* pueden ser la portabilidad y con el tiempo de descarga. El tiempo de descarga se encuentra influenciado por varios factores entre ellos: el equipo cliente, su velocidad de Internet (que puede variar constantemente) e incluso la cantidad de clientes conectados al servidor. Por otro lado, la cantidad de datos en el archivo de descarga influye en el tiempo de descarga (no al revés), también afecta el rendimiento del equipo ya que si el tamaño del archivo es muy grande puede consumir toda la memoria del equipo. Tomando en cuenta esto se decidió incorporar la cantidad adecuada de datos al grupo de dimensiones seleccionadas. En nuestro estudio a esta dimensión se le llama *Volumen de datos* y en los gráficos de resultados aparece como *fácil de manejar*. Se le llama fácil de manejar por que la publicación de archivos demasiados grandes imposibilitan la descarga y manipulación de los datos, mientras que si la cantidad de datos es adecuada es fácil de manejar.

La *pertinencia* según Naumann y Rolker [69] es el “grado en que la información satisface las necesidades de los usuarios”. En el contexto de este trabajo se puede entender como el grado en que un conjunto de datos satisface la necesidad del usuario. Esta característica tiene una relación directa con la necesidad del usuario y por lo tanto variará de acuerdo a cada usuario/necesidad. En este sentido, quien publica conjuntos de datos debe identificar y conocer las necesidades de sus usuarios. Loshin [56] dice que un atributo no relevante es aquel que, si se eliminara, no tendría ningún efecto en la aplicación que se está ejecutando actualmente ni en ninguna de las características futuras de dicha aplicación. Considerando lo anterior y aplicándolo al contexto de este trabajo se puede decir que un conjunto de datos no es relevante si al dejarse de publicar no tiene efecto en la satisfacción de las necesidades del grupo de usuarios para los que se publican los conjuntos de datos. Esto nos permite establecer una relación entre la relevancia/importancia de un conjunto de datos y las necesidades de los usuarios. Observamos que las características relevancia e importancia están relacionadas a la pertinencia. La pertinencia es la octava dimensión considerada en nuestro estudio.

Respecto a las dimensiones relacionadas con el tiempo Loshin [56] considera Currency y Timeliness. La primera se refiere “al grado en que la información está al día con el mundo que modela”. Es decir, la rapidez con la que se actualizan los datos. La segunda se refiere a “la expectativa de tiempo para el acceso a la información, y se puede medir como el tiempo entre el momento en que se espera la información y cuando está disponible para su uso”. Batini y Scannapieco [8] consideran una tercera dimensión relacionada con el tiempo denominada Volatility, esta se caracteriza por la frecuencia con la que los datos varían en el tiempo. La dimensión de tiempo es la novena que consideramos para evaluar los datos abiertos y consiste en medir si los datos están *actualizados*.

Larry English [37] se refiere a la *claridad contextual* como el grado relativo en que la presentación de datos permite al usuario comprender el significado de los datos y evitar malas interpretaciones. Esto incluye la intuición de la disposición de los datos junto con etiquetas y otra información descriptiva para facilitar la comunicación. Los datos abiertos por sí solos sin un contexto que les de sentido no son útiles. Por ello incluimos la *claridad contextual* como la décima de las dimensiones de datos abiertos incluidas en nuestro trabajo.

El tema de *licenciamiento* de los datos abiertos está en sus mismos principios, Opengovdata [74] lo indica en su punto número ocho. El esquema propuesto por Berners-Lee [10] ofrece una estrella si los datos tienen licencia abierta. El Open Data White Paper [61] publicado por Reino Unido señala en su tercer punto “los datos públicos se publicarán bajo la misma licencia abierta que permite la reutilización gratuita, incluida la reutilización comercial”. Por lo anterior se incluye el licenciamiento como undécima dimensión para evaluar los datos abiertos en este trabajo.

El quinto punto de el Open Data White Paper [61] dice “los datos públicos se publicarán utilizando estándares abiertos y siguiendo las recomendaciones pertinentes del World Wide Web Consortium (W3C)”. Berners-Lee [10] propone que quienes entreguen datos estructurados procesables por máquina y en un formato libre merecen tres estrellas en su clasificación. La tabla 2.1 muestra los formatos más utilizados en la publicación y cuales de ellos son libres. Sin duda, el *formato* es un aspecto importante a considerar en los datos abiertos, por ello se considera dentro del conjunto de dimensiones a evaluar en este estudio.

Es necesario que los conjuntos de datos publicados puedan ser interpretados de la forma correcta por el usuario, en caso contrario es casi mejor no publicarlos. Para que el usuario pueda comprender lo que se publica es necesario que los conjuntos estén bien descritos (forma general) y que posean la documentación (forma específica) que permita tener claros los detalles de las publicaciones. Según Batini y Scannapieco [8] la interpretabilidad se refiere a la “documentación y los metadatos que están disponibles para interpretar correctamente el significado y las propiedades de las fuentes de dato”. Por otro lado, el estándar ISO/IEC 25012 define comprensibilidad como el “grado en el que los datos tienen atributos que permiten ser leídos e interpretados por los usuarios y son expresados utilizando lenguajes, símbolos y unidades apropiados en un contexto de uso específico”. Naumann y Rolker [69] indican que la documentación es la cantidad y utilidad de documentos con metainformación.

Para que el usuario de datos abiertos puedan comprender e interpretar bien los datos es necesario una *descripción* que le oriente de forma general y de ser necesario la *documentación* respectiva le brinde detalles sobre los datos. Por ello, consideramos que la interpretabilidad

y la comprensibilidad dependen de la descripción y la documentación. Con estas dos dimensiones (descripción y la documentación) incluimos catorce dimensiones al conjunto.

Como ya se comentó la interoperabilidad no es una característica de los datos. Por tanto, no se puede evaluar en los conjuntos de datos. Sin embargo, el cuarto principio de la carta internacional de datos abiertos [73] dice “comparables e interoperables”. También el numeral *H* que describe los datos abiertos de investigación propuestos por la OCDE [39] hace referencia a la interoperabilidad. Dada que la interoperabilidad tiene muchos componentes y que no son abordables en este trabajo nos limitaremos a considerar la estandarización de los datos que ya ha sido comentada y los metadatos. Los *metadatos* también guardan alguna relación con las características descripción y documentación, pues los metadatos comprenden la información disponibles que describe el concepto, el contenido de los archivos, los procesos aplicados a los archivos, entre otras cosas. El Open Data White Paper [61] dice “los organismos públicos deberían publicar metadatos relevantes sobre sus conjuntos de datos... y deben publicar descripciones de apoyo del formato, procedencia y significado de los datos”. La norma técnica para publicación de datos abiertos en Chile [32] define un conjunto de metadatos mínimos que deben poseer los conjuntos publicados. (Para detalles ver la figura 4.2). Un grupo de Chief Data Officers (CDOs) denominados red cívica de analistas ofrecen ocho pautas que buscan mejorar las capacidades de los portales de datos del gobierno. Entre ellas mejorar la gestión y el uso de metadatos. Por lo anterior incluimos a los *metadatos* como la décimo quinta dimensión agregada a esta propuesta.

El Índice de Transparencia Activa en Línea [51] propone entre sus indicadores que el sistema provea un “servicio de búsqueda”. En este sentido, el Centro de Investigación y Docencia Económicas [38] propone en su trabajo “Métricas de transparencia” dos dimensiones relacionadas con las búsquedas. Primero, “Indexable por los motores de búsqueda”, que como su nombre lo indica se relaciona con los resultados en los buscadores Web. Segundo, “Motor de búsqueda” que indica que los portales deben de incluir un motor de búsqueda interno, así como la posibilidad de realizar búsquedas de contenidos utilizando filtros por fecha, palabra clave o tema. Por otro lado, la norma técnica para publicación de datos abiertos en Chile [32] define un catálogo como “la entrada organizada dirigida al usuario de los datasets publicados por el organismo, dirigido para facilitar el acceso simple, abierto, y sin restricciones de uso de los datos que posee”. La norma también indica que el catálogo debe poseer un listado completo de los conjuntos de datos del portal. Además el catálogo debe permitir la navegación, búsqueda y consulta simple de los datos contenidos en el repositorio. Para esto debe contar con todas las funcionalidades necesarias (categorizaciones, navegación, búsqueda de texto, lenguaje de consulta, etc.). La décimo sexta dimensión agregada a esta propuesta fue *Catálogo de datos*.

Naumann y Rolker [69] indican que la atención al cliente es la “cantidad y utilidad del soporte en línea a través de texto, correo electrónico, teléfono, etc.”. En el contexto de datos abierto el Índice de Transparencia Activa en Línea [51] propone entre sus indicadores que el sistema ofrezca un “atención en línea”. En este trabajo se le denomina a la dimensión relacionada al soporte y atención en línea *contactable*.

El estándar ISO/IEC 25012 define trazabilidad como el “grado en el que los datos tienen atributos que proporcionan un camino de acceso auditado a los datos o cualquier otro cambio realizado sobre los datos en un contexto de uso específico”. En los datos abiertos

Nombre	Motivada por	Definición
Complejidad	[37]	
Procesable por máquina	[10], [61]	
Primario	[56]	
Estandarizado	[56]	
Portabilidad	ISO/IEC 25012	
Capacidad de relacionarse	[34]	
Volumen de datos	[100]	
Pertinencia	[69]	El grado en que un conjunto de datos satisface la necesidad del usuario
Actualizado	[56]	
Claridad contextual	[37]	
Licencia	[74]	
Formato abierto	[61]	
Documentación	[8], [69]	Evalúa la existencia de un escrito amplio que permite comprender de forma detallada el contenido del conjunto de datos publicado
Descripción	[8], [69]	Evalúa la existencia de un párrafo de texto que caracterice lo que contiene el conjunto de datos publicados
Metadatos	[32], [73]	
Catálogo de datos	[32], [38], [51]	
Contactable	[69]	
Verificable/trazable	ISO/IEC 25012	
Consistencia	[100], [69]	
Coherencia	[25]	Evalúa la relación lógica entre el contenido del conjunto de datos, considerando principalmente los nombre de los campos y las unidades de medida utilizadas
Sostenibilidad	[39]	Evalúa que la publicación de los conjuntos de datos sea de forma periódica, sin retrasos
Privacidad	[56], [39]	
Seguridad	[100],[69], [39]	
Almacenamiento/custodia	[39]	Evalúa si los metadatos indica que organización es la responsable de la custodia permanente y preservación de los datos

Tabla 3.1: Esta tabla resume las dimensiones propuestas en este trabajo, considerando la dimensión que motivo la definición y creación de la misma

Fuente: Creación propia

también es necesario conocer los detalles de creación y modificación de los conjuntos de datos, por lo que *trazabilidad* se agregó como décimo octava dimensión.

Wang y Strong [100] dicen que la consistencia de representación es “la medida en que los datos siempre se presentan en el mismo formato y son compatibles con los datos anteriores”. Para Naumann y Rolker [69] es el “grado en que la estructura de la información se ajusta a la de otras fuentes”. En este trabajo de datos abiertos se pretende medir que los conjuntos de datos publicados en forma periódica respeten el formato y la estructura. Es decir, que los campos tenga el mismo tipo, tamaño, nombre y orden. En este trabajo a esta dimensión se le llama *consistencia*.

La *coherencia* según el Handbook on Constructing Composite Indicators: Methodology and User Guide de la OCDE [25] es:

La coherencia de los productos de datos refleja el grado en que están conectados lógicamente y son mutuamente consistentes, es decir, la idoneidad de los datos para combinarse de forma fiable de diferentes maneras y para diversos usos. La coherencia implica que el mismo término no debe usarse sin explicación para diferentes conceptos o elementos de datos; que no se deben usar términos diferentes para el mismo concepto o elemento de datos sin explicación; y que las variaciones en la metodología que puedan afectar los valores de los datos no deberían hacerse sin explicación.

En el contexto de los datos abiertos también se requiere que los conjuntos sean coherentes, en especial cuidado con medidas y valores. Que además de estandarización haya coherencia que permita claridad y fácil entendimiento.

La dimensión *sostenibilidad* se plantea directamente relacionada con el uso de datos en forma periódica. Para la OCDE [39] la sostenibilidad significa “tomar la responsabilidad administrativa de las medidas para garantizar el acceso permanente a los datos que se ha determinado que requieren retención a largo plazo”. Las dimensiones contempladas en este estudio relacionadas a la sostenibilidad de la publicación de conjuntos de datos son la sostenibilidad periódica y la sostenibilidad permanente. La primera consiste en que los datos se deben publicar de forma periódica según la ley o previo acuerdo, esta publicación debe garantizarse de forma sostenible. Es decir, sin que se incumpla provocando retrasos respecto al periodo establecido. La segunda forma se refiere al a custodia y almacenamiento de forma permanente a través del tiempo. Es decir, que nunca deje de estar en línea o que sea preservada por alguna entidad y que sea accesible. En este trabajo se denomina a la sostenibilidad periódica simplemente *sostenibilidad* y a la sostenibilidad permanente *custodia/almacenamiento*.

Según Loshin [56] la privacidad “implica la forma en que los usuarios no autorizados no pueden acceder a los datos y garantiza que los datos estén protegidos contra la visualización no autorizada”. Por otro lado, la OCDE [39] ofrece un visión más cerca al contexto de datos abiertos y dice sobre la privacidad y confidencialidad “los datos sobre sujetos humanos y otros datos personales están sujetos a acceso restringido según las leyes y políticas nacionales para proteger la confidencialidad y la privacidad. Sin embargo, los custodios de dichos datos deben considerar los procedimientos de anonimización o confidencialidad que garanticen un nivel

satisfactorio de confidencialidad para preservar la mayor utilidad posible de los datos...”. En este sentido, la legislación de cada país se sobrepone ante cualquier definición y los conjuntos de datos publicados deben estar sujetos a la ley del país donde se publican. La privacidad es la vigésima tercera dimensión adoptada en este trabajo.

Para Wang y Strong [100] la seguridad de acceso es “la medida en que el acceso a los datos puede restringirse y por lo tanto, se mantienen seguros”. Mientras que para Naumann y Rolker [69] es el “grado en que la información se pasa de forma privada del usuario a la fuente de información y viceversa”. Por otro lado, para la OCDE [39] “la seguridad, con respecto a los datos, metadatos y descripciones pertinentes, deben estar protegidos contra pérdidas intencionales o no intencionales, destrucción, modificación y acceso no autorizado de conformidad con protocolos de seguridad explícitos. Los conjuntos de datos y el equipo en el que se almacenan deben protegerse también de los riesgos ambientales como el calor, el polvo, las sobretensiones eléctricas, el magnetismo y las descargas electrostáticas.” Desde la perspectiva de los datos abiertos la dimensión de seguridad envuelve todo lo mencionado. Sin embargo, desde el punto de vista del usuario de los datos solo puede verificar si los accesos son autorizados. Consideraremos la *seguridad* como parte del conjunto de dimensiones para datos abiertos que se plantean en este trabajo.

La figura 3.19 muestra un cuadro dividido en cuatro partes. Cada parte se encuentra subdividida en dos y corresponden al usuario técnico y no técnico respectivamente. De izquierda a derecha y de arriba hacia abajo, el primer cuadrante presenta las características que corresponden a los usuarios *productores* que usan los datos en forma estática, o sea, para crear productos a corto plazo. El segundo cuadrante presenta las características que corresponden a los usuarios *consumidores* que usan los datos en forma estática, en otras palabras, a corto plazo. El tercer cuadrante presenta las características que corresponden a los usuarios *productores* que usan los datos en forma dinámica, o sea, para crear productos a largo plazo. El cuarto cuadrante presenta las características que corresponden a los usuarios *consumidores* que usan los datos en forma dinámica, es decir, a largo plazo.

Se puede entender que todo usuario productor es al mismo tiempo consumidor, por ello la similitud de las características. Sin embargo, para los usuarios productores la característica primaria es fundamental ya que permite una mayor reutilización. Las características general y específicas por uso y usuario se muestran en la 3.20.

Para los usuarios técnicos algunos requisitos claves que deben tener los datos son: que sean procesables por máquina, y que exista una buena descripción y documentación que permita comprender la estructura organización y relación entre los datos. La integridad y la seguridad de acceso a los datos contribuye a la confiabilidad. La certeza de que los datos estén completos es un factor relevante para técnicos, como para los no técnicos, por ejemplo, si se desea tener los datos de salud de todos los niños entre cinco y diez años, pero hay un alto porcentaje de registros que tienen nulo el campo de fecha de nacimiento, probablemente la estadística será incorrecta.

Para los usuarios consumidores no técnicos, como los ciudadanos, las dimensiones más interesantes son: contar con una guía o catálogo de datos que le permita saber qué datos existen, la facilidad de entender e interpretar los datos, es decir, que sean comprensibles; que esté claro como se obtuvieron/crearon los datos, en qué contexto, con qué se relacionan en

Uso/Usuario	Técnico	No técnico	Técnico	No técnico
C o r t o	Productor de datos		Consumidor de datos	
	Completos	Completos	Completos	Completos
	Procesable por máquina	Primarios	Procesable por máquina	Primarios
	Primarios	Estandarizados	Primarios	Estandarizados
	Estandarizados	Fácil de manejar	Estandarizados	Fácil de manejar
	Portabilidad	Actualizado, Oportuno	Portabilidad	Actualizado, Oportuno
	Capacidad de relacionarse	Claridad contextual	Capacidad de relacionarse	Claridad contextual
	Claridad contextual	Licenciamiento	Claridad contextual	Licenciamiento
	Formato abierto	Formato abierto	Formato abierto	Formato abierto
	Documentación	Documentación	Documentación	Documentación
	Descripción	Descripción	Descripción	Descripción
	Metadatos	Catálogo de datos	Metadatos	Catálogo de datos
	Contactable	Metadatos	Contactable	Metadatos
	Seguridad	Contactable	Seguridad	Contactable
	Coherencia	Verificable /trazable	Coherencia	Verificable /trazable
		Coherencia		Coherencia
	Privacidad		Privacidad	
	Seguridad		Seguridad	
	Almacenamiento/custodia		Almacenamiento/custodia	
L a r g o	Completos	Completos	Completos	Completos
	Procesable por máquina	Primarios	Procesable por máquina	Primarios
	Primarios	Estandarizados	Primarios	Estandarizados
	Estandarizados	Fácil de manejar	Estandarizados	Fácil de manejar
	Portabilidad	Actualizado, Oportuno	Portabilidad	Actualizado, Oportuno
	Capacidad de relacionarse	Claridad contextual	Capacidad de relacionarse	Claridad contextual
	Claridad contextual	Licenciamiento	Claridad contextual	Licenciamiento
	Formato abierto	Formato abierto	Formato abierto	Formato abierto
	Documentación	Documentación	Documentación	Documentación
	Descripción	Descripción	Descripción	Descripción
	Metadatos	Catálogo de datos	Metadatos	Catálogo de datos
	Contactable	Metadatos	Contactable	Metadatos
	Consistencia	Contactable	Consistencia	Contactable
	Coherencia	Verificable /trazable	Coherencia	Verificable /trazable
	Sostenibilidad	Consistencia	Sostenibilidad	Consistencia
	Seguridad	Coherencia	Seguridad	Coherencia
	Sostenibilidad		Sostenibilidad	
	Privacidad		Privacidad	
	Seguridad		Seguridad	
	Almacenamiento/custodia		Almacenamiento/custodia	

Figura 3.19: Distribución de las dimensiones de acuerdo a sus usos y usuarios.

Fuente: Elaboración propia

el mundo real, lo que se conoce como claridad contextual. Cada usuario tendrá su interés propio, por lo que la importancia y relevancia de los datos será subjetiva; para algunos los datos deberán estar actualizados, otros necesitarán de un histórico de datos, así la pertinencia de los datos dependerá del usuario y su uso específico.

Los usuarios productores no técnicos comparten algunos intereses con los consumidores no técnicos en lo que a las propiedades de datos se refiere, por ejemplo, que se encuentren

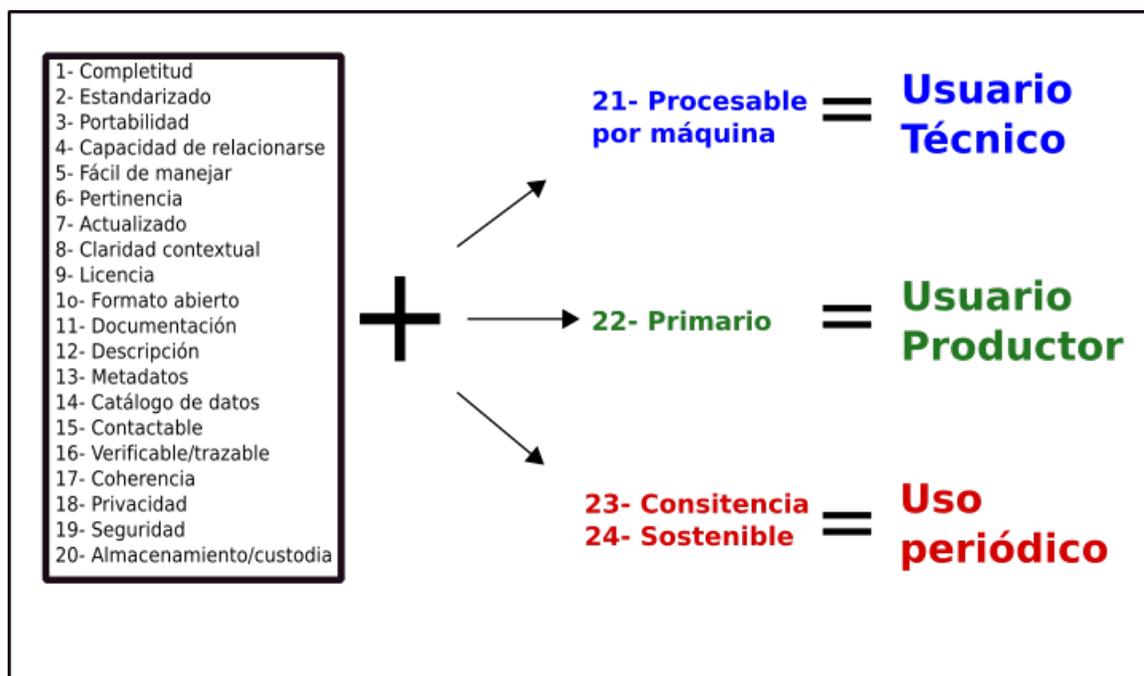


Figura 3.20: En el primer cuadro a la izquierda se muestra un conjunto de dimensiones necesarias para todos los usuarios, posteriormente se muestra las dimensiones relacionadas a usos/usuarios específicos

Fuente: Elaboración propia

en forma granular o primario. La actualización que también dependerá de la finalidad de los datos. Para el caso de los productores no técnicos tanto dinámicos como estáticos hay unas cualidades de los datos claves que les interesa, por ejemplo, las que se relacionan con el área legal, como la privacidad y la licencia de uso, que los datos sean verificables y trazables para garantizar su credibilidad.

La sostenibilidad a través del tiempo es fundamental para un usuario productor que requiera de un uso dinámico, o sea, aquel que requiera hacer la operación de extracción y manipulación de datos de forma periódica (en este caso los datos actualizados). También que se garantice la preservación de los datos para que su fuente sea validable a través del tiempo y el apoyo legal respectivo. Otro aspecto importante en la posibilidad de establecer contacto con el responsable de los datos, esta propiedad es llamada contactable o soporte al cliente en otros contextos.

Por otro lado, la licencia abierta para que los usuarios puedan usar, modificar, distribuir, etc. y el formato abierto del archivo donde se almacena y transporta el conjunto de datos son dimensiones que se relacionan con *el nivel de apertura de los datos abiertos*.

3.5.1. Descripción de las dimensiones

A continuación se describen las dimensiones que serán utilizadas en la metodología que apoyará la evaluación del nivel de apertura, capacidad de reutilización y accesibilidad de los

datos abiertos:

- **Completo:** Una fuente de datos está completo si todos los registros tienen todos los valores diferentes de null, es decir, todos los campos de todos los registros tienen su valor correspondiente del mundo real.
- **Procesable por máquina:** Una fuente de datos es procesable por máquina si posee la estructura y el formato idóneo para ser manipulado por medio de un lenguaje de programación.
- **Primario:** Que un dato sea primario tiene que ver con la precisión en la recolección de su valor, un dato es primario si su granularidad es alta. Por ejemplo, la fecha (momento) de nacimiento dividida por día, mes y año. Quizá más granular, con horas, minutos y segundos (... nanosegundos).
- **Licenciamiento:** En esta dimensión hay dos aspectos a evaluar, la primera es la autorización legal para acceder y utilizar los datos. La segunda se corresponde con la autorización legal de utilizar los programas para acceder a los datos, se encuentra directamente relacionado al formato del archivo en el que están almacenados los datos.
- **Consistencia:** Evalúa que el tipo, tamaño de los datos siempre sean los mismos para un campo determinado y que el nombre del campo tampoco varíe. En un mismo conjunto de datos.
- **Coherencia:** También considera que en fuentes de datos relacionadas la cualidad de un mismo objeto del mundo real tenga el mismo nombre. Por ejemplo, si en una fuente de datos aparece el campo rut para la entidad persona y en otra fuente relacionada aparece cédula para la misma entidad persona hay incoherencia.
- **Metadatos:** Esta dimensión evalúa si los datos incorporan metadatos, al menos de los tipos: descriptivos, administrativos y estructurales.
- **Documentación/Descripción:** Dado que en muchas fuentes de datos se utilizan abreviaturas o algunas palabras técnicas es necesario incluir mínimo una descripción con el nombre completo de las variables o un diccionario de datos que permita conocer de que se trata cada campo en la fuente de datos.
- **Privacidad:** Mide si la fuente de datos respeta la privacidad según lo establecido en la legislación del país.
- **Estandarizado:** Esta propiedad de los datos reconoce si los valores de los datos se encuentran dados en unidades establecidas por el sistema internacional de unidades, que permita la interoperabilidad.
- **Sostenible:** Tiene que ver con el compromiso que ofrece el proveedor de los datos de suministrarlos en forma periódica sin falta.
- **Seguridad:** El lugar donde residen los datos debe poseer las condiciones de seguridad necesarias para garantizar el funcionamiento continuo del servicio y para evitar que los datos sean alterados sin el permiso correspondiente.
- **Portabilidad:** Mide la capacidad de movilizar/compartir/descargar los datos a otros sistemas conservando las calidades de los datos.
- **Capacidad de relacionarse:** En este contexto quiere decir que la fuente de datos debe tener los identificadores únicos de recursos (URI) y las capacidades para relacionarse con otras fuentes de datos y contribuir a la interoperabilidad.

- **Pertinencia:** Mide la utilidad que tenga el dato para satisfacer el propósito por el cual el consumidor necesitaba el dato.
- **Actualizado:** Los valores en el mundo real cambian constantemente, esta propiedad evalúa que los datos de la fuente estén al día, es decir, que también cambien conforme cambia el mundo real.
- **Claridad contextual:** Los datos por sí solos no comunican nada, no se pueden interpretar con exactitud si no se conoce de donde se toman y a qué hacen referencia en el mundo real. Claridad contextual evalúa que esté claro a qué se refieren en el mundo real los datos.
- **Contactable:** Evalúa si en la página donde se publica la fuente de datos existe un mecanismo que permita establecer comunicación con el proveedor de los datos, de forma que brinde soporte al usuario.
- **Catálogo de datos:** Evalúa si existe una guía o catálogo de datos que oriente al usuario respecto a los conjuntos de datos que se pueden encontrar en el portal. Esta puede consistir de una lista de conjuntos de datos ordenados por temas y con posibilidades de filtrar por temas u otra característica. Puede ser pertinente la posibilidad de realizar búsquedas.
- **Verificable/Trazabilidad:** Mide la capacidad de los datos de ofrecer atributos, que faciliten conocer las operaciones de auditoría, la creación y modificaciones realizadas, quién las realizó y con qué objetivo.
- **Volumen de datos:** Esta dimensión mide el tamaño del archivo donde se encuentra almacenado el conjunto de datos. Definir un tamaño ideal para un archivo de este tipo no es tarea fácil, y no está dentro de nuestros objetivos. Podemos establecer como precedente la necesidad de definir un tamaño estándar para que posteriormente se analice. De forma intuitiva podemos reconocer cuatro alcances: un archivo de datos que se pueda descargar a una velocidad promedio puede tener un tamaño inferior a doscientos cincuenta megabytes, esto es fácilmente manipulable. Segundo, un archivo de hasta un gigabyte se puede manipular de manera manejable. Tercero, un archivo de datos de un gigabyte a cuatro gigabytes es un archivo difícil de manejar. Mientras que archivos superiores de 4 gigabytes, en términos de terabytes y petabytes se pueden considerar intratables.
- **Almacenamiento/Custodia:** Esta dimensión guarda relación con la necesidad de mantener los datos almacenados por largo tiempo, evalúa si los metadatos indica que organización es la responsable de la custodia permanente y preservación de los datos.

Capítulo 4

Métricas e indicadores para datos abiertos

En la primera sección de este capítulo se estudian los trabajos que proponen métricas y realizan mediciones relacionados con la transparencia y la publicación de la información en los portales de transparencia de los gobiernos. En la segunda sección se proponen un conjunto de métricas para medir la apertura, capacidad de reutilización y accesibilidad de los conjuntos de datos abiertos de gobierno. Estas métricas surgen a partir de las dimensiones estudiadas en la sección [3.5.1](#).

4.1. Revisión de métricas de información de transparencia

Al hacer una revisión de los trabajos realizados para evaluar los portales que publican datos abiertos se ha encontrado que estos se enfocan mayormente en la medición de la *información* de transparencia y en la capacidad de los portales para ofrecer dicha información. A nivel global encontramos los trabajos realizados por Transparencia Internacional, como el Índice de percepción de la corrupción, el barómetro global de corrupción, entre otros. Estos no miden portales si no que ofrecen los resultados de una recopilación de estudios en los que se entrevista a expertos y otras fuentes para generar un índice.

Más localmente en Chile el Consejo para la Transparencia publicó en el 2013 el Índice de Transparencia y Acceso a la Información (ITAI) [\[33\]](#) el cual está conformado por tres dimensiones:

1. La primera es la dimensión *uso* que contempla:
 - (a) Acciones de transparencia activa
 - (b) Acciones de transparencia pasiva

2. La segunda dimensión es *ciudadana* y se subdivide en:

- (a) Valoración
- (b) Conocimiento
- (c) Acceso

3. La última dimensión es *institucional* y comprende las sub dimensiones:

- (a) Valoración
- (b) Conocimiento
- (c) Capacidad

La figura 4.1 presenta las preguntas/indicadores que corresponden a la dimensión institucional y a la subdimensión conocimiento propuestos en ITAI [33]. Este cuadro evalúa aspectos sobre el conocimiento que los ciudadanos tienen sobre la ley y el derecho de acceso a la información.

También en el 2013, en ese mismo país se creó la “Norma técnica para publicación de datos abiertos en Chile” [32]. Este documento no muestra indicadores ni métricas, pero establece una serie de lineamientos a cumplir, que pueden transformarse para medir el cumplimiento de algunas dimensiones de la publicación de datos. Define los formatos para: imágenes, datos geográficos, datos estadísticos. Con respecto a la publicación establece: los metadatos obligatorios y los recomendados, la licencia, el tiempo de actualización y la estructura del catálogo. Estos lineamientos tienen como base algunos aspectos técnicos generales, y parte de la legislación nacional de Chile.

La figura 4.2 muestra los nombres y descripciones de los metadatos obligatorios propuestos en la Norma técnica para publicación de datos abiertos en Chile [32]. Este es un ejemplo de lo que se puede adecuar para crear indicadores.

En Costa Rica algunas organizaciones también han propuestos métodos de evaluación, por ejemplo el Centro de Investigación y Capacitación en Administración Pública de la Universidad de Costa Rica (en adelante CICAP) y el Gobierno Digital de Costa Rica crearon el “Documento Metodológico del Índice de Transparencia del Sector Público Costarricense, basado en Sitios Web” [102], el cual propone una estructura compuesta por dimensiones, variables sub variables e indicadores. Las cuatro dimensiones son: acceso a la información, rendición de cuentas, participación ciudadana, datos abiertos de gobierno. A continuación se mencionan las variables por cada dimensión.

La primera dimensión es *acceso a la información* y posee las siguientes cinco variables:

- 1. Información institucional
- 2. Información del personal institucional

Sub Dimensión	Ítems de Mediciones	Variable Observada	Alternativas de respuesta	Responde
Valoración	Valoración	P10.1 Toda la información que poseen los organismos públicos, debe ser pública	Muy de acuerdo / De acuerdo	Todos
		P28 ¿Usted ha solicitado información a un organismo público en los últimos 12 meses?	Si	Todos
	Reconocimiento intuitivo del Derecho de Acceso a la Información	P12.1 Cualquier persona tiene derecho a acceder a información de cualquier organismo público	Muy de acuerdo / De acuerdo	Todos
		P12.2 Cuando una persona solicita información a un organismo público, este está obligado a entregarla	Muy de acuerdo / De acuerdo	Todos
		P12.3 Si un organismo público no entrega la información solicitada, existe derecho a reclamar	Muy de acuerdo / De acuerdo	Todos

Figura 4.1: Indicadores/preguntas creados para la dimensión ciudadano y la sub dimensión valoración, en ITAI

Fuente: Índice de transparencia activa y acceso a la información [33], Chile

3. Procesos de contratación
4. Servicios y procesos institucionales
5. Herramientas y condiciones

La segunda es *rendición de cuentas* e incorpora las cuatro variables que se muestran a continuación:

1. Finanzas públicas
2. Planes y cumplimiento
3. Informes institucionales y de personal
4. Toma de decisiones

La tercer dimensión considerada es *participación ciudadana* y sus variables se presentan en seguida:

1. Invitación a la participación
2. Mecanismos de participación
3. Calidad de los mecanismos de participación

Por último, la dimensión *datos abiertos de gobierno* considera entre sus variables a:

#	Nombre	Descripción
1	Título	Nombre definido para el dataset.
2	Descripción	Resumen descriptivo del dataset, especificando el contenido de recurso.
3	Institución	Nombre de la institución o entidad responsable de hacer el dataset públicamente disponible.
4	Licencia	Información sobre los derechos de uso del dataset.
5	Categoría	Clasificación conceptual básica del dataset en sistemas de categorías o taxonomías disponibles.
6	Fecha de publicación	Punto de tiempo asociado al evento de publicación del dataset.
7	Recurso	Contenido y características físicas de un dataset. Un mismo dataset puede tener varios recursos, siempre y cuando se busque publicar la misma información utilizando distintos formatos (por ejemplo, en formato XML, en formato comprimido, en formato CSV, etc). A su vez, cada recurso debe contener los siguientes metadatos: enlace de distribución, tamaño, formatos, descripción, fecha de los metadatos.

Figura 4.2: Nombres y descripciones de los metadatos obligatorios, propuestos en la Norma técnica para publicación de datos abiertos en Chile

Fuente: Norma técnica para publicación de datos abiertos en Chile [32]

1. Presupuesto público
2. Ejecución de presupuesto
3. Contratación administrativa
4. Estadísticas o registros del sector

Entre los anexos del trabajo realizado por el CICAP se encuentran algunos cuadros con las subvariables e indicadores. Algunos de estos están relacionados con los datos, de manera que evalúan características técnicas de los datos. En la figura 4.3 se muestra las subvariables e indicadores para la variable “ejecución del presupuesto” creados para la dimensión datos abiertos de gobierno, además de la fuente de donde se toma la idea para crear el indicador y el puntaje asignado en el modelo de evaluación.

En Uruguay también existe una herramienta para medir la transparencia, denominado Índice de Transparencia Activa en Línea (ITAeL) creado por el Centro de Archivos y Acceso a la Información Pública (CAInfo) y la Universidad Católica del Uruguay. Este instrumento posee dos dimensiones: la primera de ellas denominada información obligatoria y la segunda interacción y participación. Estas dimensiones se encuentran relacionadas con el contenido y forma en que se estructura la información y con la interacción que realizan los usuario con esta.

Para medir la dimensión de información obligatoria los autores del ITAeL proponen los indicadores que cito a continuación textualmente:

1. Creación del organismo.
2. Evolución histórica.
3. Descripción de funciones, competencias, cometidos, del organismo.

Sub-variables e Indicadores para "Ejecución del Presupuesto"				
4.2.1	Apertura técnica	Evalúa la disponibilidad técnica del dato, si está en formato digital, si está disponible en línea, si tiene un formato legible por computadoras, si está disponible en masa y si está actualizado.	Open Knowledge (2015)	0.5400
4.2.1.4	Legible por computadora	Los datos tienen un formato que puede ser fácilmente estructurado por una computadora (los datos pueden ser digitales, pero no legibles a máquina, por ejemplo PDF y HTML) Si/No	Open Knowledge (2015)	0.4400
4.2.1.5	Disponible en masa	Los datos pueden ser descargados en su conjunto. No estarían disponibles en masa si se ofrecen sólo partes del conjunto de datos. Si/No	Open Knowledge (2015)	0.2800
4.2.1.6	Datos actualizados	Los datos están actualizados y disponibles de manera oportuna en su espacio temporal. Si/No	Open Knowledge (2015)	0.2800
4.2.2	Apertura legal	Evalúa la disponibilidad legal del dato, si está disponible al público general, si está libre de costos, y si tiene una licencia expresa que permita su libre uso, reuso y redistribución.	Open Knowledge (2015)	0.4600
4.2.2.3	Licencia abierta	Los datos tienen que indicar los términos de uso o licencia que permiten que cualquiera pueda utilizarlos libremente, pueda reutilizarlos o redistribuirlos. La licencia debe estar disponible. Si/No	Open Knowledge (2015)	1.0000

Figura 4.3: Sub variables e indicadores para la variable “ejecución del presupuesto”
Fuente: Documento Metodológico del Índice de Transparencia del Sector Público Costarricense, basado en Sitios Web [102], Costa Rica

4. Estructura organizativa (organigrama).
5. Marco jurídico aplicable al organismo.
6. Perfil de los diferentes puestos de trabajo a partir de jefe de dpto, gerente, director o equivalente hasta el jerarca.
7. Currículum Vitae de Autoridades.
8. Contacto: nombre de quienes ocupan los cargos a partir de jefe de departamento con teléfono, correo electrónico, dirección postal.
9. Programas operativos (de largo y corto plazo: entre 5 años y 1 año).
10. Listado con los servicios que ofrece el organismo (si corresponde).
11. Listado de los programas que administra el organismo (si corresponde).
12. Trámites para acceder a los servicios (si corresponde).
13. Nómina de los funcionarios que cumplen funciones en el mismo estando en comisión, con indicación de compensaciones recibidas con cargo a las partidas asignadas al funcionamiento del organismo.
14. Nómina de los funcionarios contratados, con compensaciones percibidas con cargo a partidas asignadas al organismo.

15. Estructura de remuneraciones por categoría escalafonaria, funciones de los cargos y sistemas de compensación.
16. Remuneración de viáticos recibidos.
17. Listado de comisiones de servicio en el exterior de los funcionarios, viáticos percibidos, razón del viaje y resultados del mismo, incluyendo a todas las personas que integres la delegación sin excepción alguna.
18. Listado de programas de capacitación, número de funcionarios capacitados así como sus evaluaciones.
19. Convocatorias a concursos de ingreso para ocupar cargos y resultados de los mismos.
20. Presupuesto Asignado (Ingresos asignados por el presupuesto nacional).
21. Ejecución del Presupuesto.
22. Auditorías: Número y tipo. Número de observaciones realizadas por rubro de auditoría sujeto a revisión. Total de aclaraciones efectuadas por el organismo.
23. Ingresos recibidos por cualquier concepto, con indicación del responsable en la recepción, administración y ejecución.
24. Reporte anual financiero (Balance).
25. Información sobre Licitaciones, Concesiones etc. (Vínculo electrónico con el sitio compras estatales).
26. Publicaciones y/o estadísticas asociadas a los fines de cada organismo.
27. Información sobre la política de seguridad y protección de datos.
28. Responsable de la información que se publica en el sitio web.
29. Dirección postal del organismo, teléfono, email de contacto.
30. Mapa del Sitio.
31. Fecha de última actualización.

Respecto a la dimensión de interacción y participación los autores consideran los 15 indicadores siguiente (cito textualmente):

1. Servicios en línea (transacciones, trámites).
2. Teléfono gratuito, solo información.
3. Redes sociales.
4. Avisos.

5. Newsletter.
6. Notificación de actualización de la página. Feedburner (u otros).
7. Foros de opinión.
8. Servicios de búsqueda en la página.
9. Ayuda en la página.
10. Preguntas frecuentes.
11. Glosario.
12. Otros idiomas.
13. Formato de impresión.
14. Reenvío de información.
15. Atención en línea.

Algunos indicadores de la dimensión de interacción y participación (DIP) y de la dimensión de información obligatoria (DIO) tiene relación con las dimensiones citadas en la páginas 59 y posteriores. Algunas asociaciones se pueden realizar de la siguiente manera: documentación/descripción con los 9, 10 y 11 de la DIP. Contactable con el 29 de DIO y con los 2 y 15 de la DIP. Actualizado con la 31 de la DIO. Finalmente valor agregado con los 6, 13 y 14 de la DIP.

En la revisión de la literatura que comprende el contexto latinoamericano se encontró un trabajo realizado en el 2013 por Ríos y Cejudo [17]. Esta investigación se aplicó en cuatro países de Centroamérica, estos fueron: Guatemala, Honduras, Nicaragua y Panamá. La metodología usada para realizar este trabajo fue propuesta en el 2010 por el centro de Investigación y Docencia Económicas (CIDE) de México.

El CIDE publicó en el 2007 “Métrica de la transparencia en México”, en el 2010 “Métrica de la Transparencia 2010” y en el 2014 “Medición de Transparencia y Acceso a la Información en México”. La última publicación [38] del CIDE presenta 5 dimensiones: Normatividad, Portales, Usuario simulado, Órganos garantes, Sujeto obligado.

Los indicadores propuestos por CIDE en el 2014 [38] fueron:

1. Público ciudadano
2. Trámite o servicio
3. Quejas y sugerencias
4. Tiempo de descarga

5. Indexable por los motores de búsqueda
6. Reporte de fallas
7. Ruta de navegación
8. Derechos de autor o propiedad del contenido
9. Legibilidad
10. Contenidos en distintas lenguas
11. Compatibilidad de navegadores
12. Plugins adicionales
13. Motor de búsqueda
14. Mapa de sitio
15. Enlaces rotos
16. Uso de enlaces únicos
17. Datos abiertos
18. Restricción de acceso a dato

La literatura revisada mostró que la mayoría de los trabajos sobre información de transparencia se encuentran estructurados por dimensiones. Entre ellas: participación ciudadana y conocimiento de los ciudadanos acerca de las leyes de transparencia; información de la institución que debe estar publicada en los sitios de transparencia y el orden que ésta debe tener; las capacidades de los portales donde se encuentra la información de transparencia. Se observó que algunos de esos trabajos involucran en parte a los datos, como la norma técnica propuesta por el gobierno de Chile, pero ninguno tiene un enfoque estricto en datos abiertos que considere las meta-dimensiones de usuarios y usos.

4.2. Indicadores para la evaluación de datos abiertos según dimensiones estudiadas

En nuestro diario vivir utilizamos una gran cantidad de objetos, entre ellos: el teléfono celular, el computador, el televisor, platos, vasos, cucharas y muchos más. Analicemos por un momento el teléfono celular, ¿qué características tiene? Comenzaremos por algunas características que causan gran controversia a la hora de elegir el dispositivo: la marca, modelo, el sistema operativo, cantidad de memoria ram, procesador. ¿Qué más? También podemos observar que tiene color de la carcasa, tamaño, etc. Considere ahora posibles valores para dichas características. Encontramos entonces que en nuestro mundo real o material existen

Marca	Modelo	Sistema operativo
Xiaomi	REDMI Y1	Android bajo MIUI 9
iPhone	8 plus	iOS 11
Samsung	Galaxy Note8	Android

Tabla 4.1: Ejemplo de conjunto de datos.

Fuente: Creación propia.

objetos o entidades de los cuales observamos características y ellas tienen valores. Para el color de la carcasa: negro, marca: Samsung. Estos valores de atributos o características son datos. Por ejemplo, de la entidad teléfono se puede tener el conjunto de datos que se muestra en la tabla 4.1.

Por tanto, el texto iPhone que se encuentra en la primera columna y en la segunda fila de la tabla 4.1 es un dato que coincide con el valor del atributo marca de la entidad teléfono. De hecho todos los valores de las filas excepto el de la primera fila son datos. El conjunto de atributos en este caso marca, modelo y sistema operativo se consideran como la estructura del conjunto de datos, también llamado encabezados o nombres de las variables. A cada fila se le denomina registro.

En la sección 3.2 se estudiaron las dimensiones de datos. Por otro lado, en la sección 3.5 se realizó un análisis sobre los usos y usuarios, y como estos se relacionan con las dimensiones de datos estudiadas. Las dimensiones que se encuentran relacionadas con los usos y usuarios estudiados se describen en el apartado 3.5.1. A continuación se proponen un conjunto de métricas para medir la apertura, capacidad de reutilización y accesibilidad de los conjuntos de datos abiertos de gobierno según las dimensiones estudiadas.

Dimensión completitud

Esta dimensión se evalúa en un conjunto de datos, y consiste en conocer si de todos los registros del esquema existe al menos un campo en blanco, es decir, que falte un dato.

- Indicador(es)
 1. *¿Posee el conjunto de datos al menos un dato ausente en sus registros?*
 - i) Posibles respuestas
 - a) *Sí*
 - b) *No*
 - ii) Procedimiento
 - *Paso 1: Ingresar a la página de datos abiertos*
 - *Paso 2: Ubicar y descargar el archivo a evaluar*
 - *Paso 3: Abrir el archivo y verificar si existe un campo vacío o nulo*

Dimensión procesable por máquina

En esta dimensión se evalúa tanto el formato como la estructura interna del archivo, en algunas ocasiones los archivos en xls u ods son preparados con filas combinadas y sin

estructura, lo que produce dificultades a la hora de procesar los datos de forma automática. Una muestra de estructura correcta se puede ver en la tabla 4.1.

- Indicador(es)

1. *¿El archivo donde se encuentran los datos tiene una extensión o formato que facilita la automatización de la manipulación de los datos?*

- i) Posibles respuestas

- a) *Sí*

- b) *No*

- ii) Procedimiento

- *Paso 1: Ingresar a la página de datos abiertos*

- *Paso 2: Ubicar y descargar el archivo a evaluar*

- *Paso 3: Identificar la extensión (csv, xml, ods) del archivo*

2. *¿Posee el conjunto de datos en su contenido interno registros estructurados?*

- i) Posibles respuestas

- a) *Sí*

- b) *No*

- ii) Procedimiento

- *Paso 1: Ingresar a la página de datos abiertos*

- *Paso 2: Ubicar y descargar el archivo a evaluar*

- *Paso 3: Abrir el archivo y verificar si el contenido interno de los los registros es estructurado*

Dimensión primario

Esta dimensión evalúa el nivel de granularidad que poseen los datos.

- Indicador(es)

1. *¿Los datos se encuentran en un grado fino de granularidad, sin cálculos ni agregaciones?*

- i) Posibles respuestas

- a) *Sí*

- b) *No*

- ii) Procedimiento

- *Paso 1: Ingresar a la página de datos abiertos*

- *Paso 2: Ubicar y descargar el archivo a evaluar*

- *Paso 3: Abrir el archivo a evaluar*

- *Paso 4: Identificar el nivel de granularidad de los datos*

Dimensión licenciamiento

Esta dimensión evalúa tres aspectos: que el sitio muestre el tipo de licencia de los datos publicados. En segundo lugar, que la licencia provista permita reutilizar los datos. Por úl-

timo, que el formato del archivo donde se almacenan los datos permita acceder a ellos sin costo/abierto.

- Indicador(es)

1. *¿El sitio indica la licencia de los datos?*
 - i) Posibles respuestas
 - a) *Sí*
 - b) *No*
 - ii) Procedimiento
 - *Paso 1: Ingresar a la página de datos abiertos*
 - *Paso 2: Buscar la leyenda que indica el tipo de licencia de los datos*
2. *¿La licencia provista permite reutilizar los datos?*
 - i) Posibles respuestas
 - a) *Sí*
 - b) *No*
 - ii) Procedimiento
 - *Paso 1: Ingresar a la página de datos abiertos*
 - *Paso 2: Buscar la leyenda que indica el tipo de licencia de los datos*
 - *Paso 3: Verificar si la licencia indicada permite reutilizar los datos*
3. *¿El formato del archivo donde se almacenan los datos posee una licencia que permite acceder a ellos sin costo/abierto?*
 - i) Posibles respuestas
 - a) *Sí*
 - b) *No*
 - ii) Procedimiento
 - *Paso 1: Ingresar a la página de datos abiertos*
 - *Paso 2: Ubicar y descargar el archivo a evaluar*
 - *Paso 3: Identificar la extensión/formato y determinar si el programa que crea archivos con esa extensión/formato es sin costo/abierto*

Dimensión consistencia

La consistencia se evalúa en un contexto donde se publican conjuntos de datos con la misma estructura en forma periódica. Por ejemplo, el conjunto de datos de salarios de empleados de planta el cual cada mes es publicado.

Nota aclaratoria: Se recomienda realizar la evaluación al menos en tres archivos de datos del mismo tema pero en período distintos. Por ejemplo, “salarios de personal de planta” considere que hay archivos de datos de enero del 2014 a enero del 2018, entonces puedes tomar datos del mes de enero del 2014 de enero del 2016 y enero del 2018.

De los indicadores que se citan a continuación, si alguno tiene como respuesta un No entonces se estima que no hay consistencia en los conjuntos de datos.

- Indicador(es)
 1. *¿El nombre de los atributos es igual en las diferentes entregas que se realizan periódicamente?*
 - i) Posibles respuestas
 - a) *Sí*
 - b) *No*
 - ii) Procedimiento
 - *Paso 1: Ingresar a la página de datos abiertos*
 - *Paso 2: Descargar los archivos a evaluar*
 - *Paso 3: Comparar el nombre de los atributos en ambos archivos para verificar si son iguales*
 2. *¿El tipo de datos es siempre el mismo? Por ejemplo, siempre es entero*
 - i) Posibles respuestas
 - a) *Sí*
 - b) *No*
 - ii) Procedimiento
 - *Paso 1: Ingresar a la página de datos abiertos*
 - *Paso 2: Descargar los archivos a evaluar*
 - *Paso 2: Comparar el tipo de los atributos en ambos archivos para verificar si son iguales*
 3. *¿El tamaño de los datos es siempre el mismo?*
 - i) Posibles respuestas
 - a) *Sí*
 - b) *No*
 - ii) Procedimiento
 - *Paso 1: Ingresar a la página de datos abiertos*
 - *Paso 2: Descargar los archivos a evaluar*
 - *Paso 2: Comparar el tamaño de los atributos en ambos archivos para verificar si son iguales, o al menos si se considera un tamaño máximo*
 4. *¿La estructura de datos es siempre la misma?*
 - i) Posibles respuestas
 - a) *Sí*
 - b) *No*
 - ii) Procedimiento
 - *Paso 1: Ingresar a la página de datos abiertos*
 - *Paso 2: Descargar los archivos a evaluar*
 - *Paso 2: Comparar el tipo de los atributos en ambos archivos para verificar si son iguales*

Dimensión coherencia

La coherencia mide que los datos sean razonables, para comprenderlo mejor se ilustra con un ejemplo. Cuando los valores de un conjunto de datos no tienen la misma forma son difíciles de comparar. Considere el precio de tres productos, si el primero está dado en pesos chilenos, el segundo en dólares estadounidense, y el tercero toma como referencia la mitad del costo de otro producto, en este escenario se considera que no hay coherencia en los datos.

En un conjunto de datos los valores de los datos de un mismo campo deben estar ajustados a una medida única determinada.

- Indicador(es)

1. *¿En el conjunto de datos los valores de los datos de un mismo campo están ajustados a una medida única determinada?*

- i) Posibles respuestas

- a) *Sí*

- b) *No*

- ii) Procedimiento

- *Paso 1: Ingresar a la página de datos abiertos*

- *Paso 2: Ubicar y descargar el archivo a evaluar*

- *Paso 3: Abrir el archivo y verificar si los datos están ajustados a una medida única determinada*

Dimensión metadatos

La dimensión metadatos evalúa que un conjunto de datos contenga como mínimo cierta cantidad de datos sobre los datos que se encuentran el conjunto.

Los metadatos mínimos según la norma técnica chilena [32] son: 1 Título Nombre definido para el dataset. 2 Descripción Resumen descriptivo del dataset, especificando el contenido de recurso. 3 Institución Nombre de la institución o entidad responsable de hacer el dataset públicamente disponible. 4 Licencia Información sobre los derechos de uso del dataset. 5 Categoría Clasificación conceptual básica del dataset en sistemas de categorías o taxonomías disponibles. 6 Fecha de publicación Punto de tiempo asociado al evento de publicación del dataset. 7 Recurso Contenido y características físicas de un dataset. Un mismo dataset puede tener varios recursos, siempre y cuando se busque publicar la misma información utilizando distintos formatos (por ejemplo, en formato XML, en formato comprimido, en formato CSV, etc). A su vez, cada recurso debe contener los siguientes metadatos: enlace de distribución, tamaño, formatos, descripción, fecha de los metadatos.

Para que se considere que un conjunto de datos tenga metadatos debe cumplir con todos los metadatos mínimos.

- Indicador(es)

1. *¿El conjunto de datos posee los metadatos definidos como “metadatos mínimos”?*

- i) Posibles respuestas

- a) *Sí*

- b) *No*

ii) Procedimiento

- *Paso 1: Ingresar a la página de datos abiertos*
- *Paso 2: Ubicar el archivo de datos a evaluar*
- *Paso 3: Ubicar y descargar el archivo de metadatos a evaluar*
- *Paso 4: Abrir el archivo de metadatos y verificar la existencia de los metadatos indicados*
- *Aclaración: Los metadatos pueden venir en el mismo archivo de datos, por lo que los pasos de evaluación pueden variar*

Dimensión documentación

En la estructura de un conjunto de datos algunas veces el nombre de un atributo se encuentra dado por siglas o acrónimos, también puede ocurrir que el nombre del atributo sea poco común o parte de un lenguaje técnico, por lo que el usuario no puede entender los datos con facilidad.

- Indicador(es)

1. *¿Posee el conjunto de datos al menos un archivo con definiciones y/o explicaciones que ayuden al usuario a interpretar los datos?*

- i) Posibles respuestas

- a) *Sí*
- b) *No*

- ii) Procedimiento

- *Paso 1: Ingresar a la página de datos abiertos*
- *Paso 2: Ubicar y descargar el archivo a evaluar*
- *Paso 3: Ubicar y descargar el archivo de documentación a evaluar*
- *Paso 4: Abrir el archivo de documentación y verificar si el archivo tiene el contenido pertinente*

Dimensión descripción

Al ubicar un archivo de datos puede que no se conozca o comprenda con claridad su contenido, por ello es necesario una descripción breve que oriente al usuario sobre el contenido del archivo. Si es un archivo que se publica de forma periódica es suficiente una descripción general, pues todos los archivos deben tener sistemáticamente lo mismo. En caso contrario se debe especificar los que difieren.

- Indicador(es)

1. *¿Posee el sistema donde se publica el conjunto de datos una descripción del contenido del conjunto?*

- i) Posibles respuestas

- a) *Sí*
- b) *No*

- ii) Procedimiento

- *Paso 1: Ingresar a la página de datos abiertos*
- *Paso 2: Ubicar el archivo a evaluar*
- *Paso 3: Verificar si existe una descripción del archivo*

Dimensión privacidad

La legislación de cada país define los derechos de privacidad que tiene su población, los datos publicados deben estar sujetos a dichas restricciones.

- **Indicador(es)**

1. *¿Los datos publicados respetan la legislación vigente del país respecto al tema de privacidad?*
 - i) Posibles respuestas
 - a) *Sí*
 - b) *No*
 - ii) Procedimiento
 - *Paso 1: Ingresar a la página de datos abiertos*
 - *Paso 2: verificar si el archivo tiene el contenido pertinente*
 - *Paso 3: Ubicar y descargar el archivo a evaluar*
 - *Paso 4: Abrir el archivo y verificar si el archivo tiene el contenido de acuerdo a lo establecido en la leyes respecto a la privacidad*

Dimensión estandarizado

La estandarización mide que los datos se encuentren ajustados a una única medida y que esta medida sea exacta y reconocida mediante una norma internacional . Por ejemplo, si el tamaño de todos los productos en los conjuntos de datos está dado en Jemes (Un jeme según la Real Academia Española es la “distancia que hay desde la extremidad del dedo pulgar a la del índice, separado el uno del otro todo lo posible”) la medida no es exacta y no se encuentra en el sistema internacional de unidades. En un caso como el descrito no se considera que los datos estén estandarizados.

En un conjunto de datos los valores de los datos de un mismo campo deben estar ajustados a una medida única y esta medida debe ser exacta y pertenecer al sistema internacional de unidades.

- **Indicador(es)**

1. *¿En el conjunto de datos los valores de los datos de un mismo campo están ajustados a una medida única la cual es exacta y pertenece al sistema internacional de unidades?*
 - i) Posibles respuestas
 - a) *Sí*
 - b) *No*
 - ii) Procedimiento
 - *Paso 1: Ingresar a la página de datos abiertos*

- *Paso 2: Ubicar y descargar el archivo a evaluar*
- *Paso 3: Abrir el archivo y verificar si los datos están ajustados a una medida única la cual es exacta y pertenece al sistema internacional de unidades*

Dimensión sostenibilidad

Esta dimensión considera la disposición y el compromiso de publicar el archivo de datos de forma periódica y continua con los datos actualizados de acuerdo a la normativa o compromiso vigente.

- **Indicador(es)**
 1. *¿El conjunto de datos se publica de forma periódica?*
 - i) Posibles respuestas
 - a) *Sí*
 - b) *No*
 - ii) Procedimiento
 - *Paso 1: Ingresar a la página de datos abiertos*
 - *Paso 2: Ubicar el archivo a evaluar*
 - *Paso 3: Verificar la fecha de publicación para determinar si los datos se entregan de forma periódica*
 2. *¿Existe un compromiso de mantener la publicación del conjunto de datos en forma periódica?*
 - i) Posibles respuestas
 - a) *Sí*
 - b) *No*
 - ii) Procedimiento
 - *Paso 1: Ingresar a la página de datos abiertos*
 - *Paso 2: Ubicar el archivo a evaluar*
 - *Paso 3: Verificar la existencia de algún anuncio que demuestre el compromiso de publicar los datos de forma periódica*

Dimensión seguridad

Es la capacidad del sistema donde residen los datos para resguardar y permitir que solo los usuarios autorizados puedan agregar, modificar y/o eliminar el contenido del sitio, incluyendo los datos por supuesto.

- **Indicador(es)**
 1. *¿Posee el sistema donde se alojan los datos algún mecanismo para agregar, modificar o eliminar conjuntos de datos, sin realizar las validaciones de usuario pertinentes de manera que ponga en riesgo la calidad de los datos?*
 - i) Posibles respuestas
 - a) *Sí*

- b) *No*
- ii) Procedimiento
 - *Paso 1: Ingresar a la página de datos abiertos*
 - *Paso 2: Verificar la existencia de mecanismos que permitan manipular conjuntos de datos, sin realizar las validaciones de usuario pertinentes*

Dimensión portabilidad

Es la capacidad del sistema donde residen los datos para permitir: movilizar, compartir, descargar los datos sin alterar las características y cualidades de los mismos.

- Indicador(es)
 1. *¿El sistema donde se alojan los datos ofrece la posibilidad de compartir, descargar los datos en bulto o través de una API?*
 - i) Posibles respuestas
 - a) *Sí*
 - b) *No*
 - ii) Procedimiento
 - *Paso 1: Ingresar a la página de datos abiertos*
 - *Paso 2: Verificar la existencia de mecanismos que permitan descargar o compartir los datos*

Dimensión capacidad de relacionarse

La capacidad de relacionarse evaluará que el conjunto de datos posea identificadores de recursos uniformes (en adelante URI por sus siglas en inglés), aunque se puede esperar que todos los recursos dentro del conjunto de datos estén identificados no es necesario para esta evaluación. Con solo el hecho de que el conjunto cuente con URI se da por entendido el inicio del proceso y se le reconoce la posibilidad de relacionarse. Se espera que si la URI está presente en el sistema o en los metadatos la labor técnica esté realizada.

- Indicador(es)
 1. *¿El conjunto de datos se encuentra asociado a un identificador de recursos uniforme, ya sea a través de metadatos o en una etiqueta en el sistema?*
 - i) Posibles respuestas
 - a) *Sí*
 - b) *No*
 - ii) Procedimiento
 - *Paso 1: Ingresar a la página de datos abiertos*
 - *Paso 2: Verificar que el conjunto de datos se encuentre asociado a un identificador de recurso uniforme*

Dimensión pertinencia

La pertinencia evalúa si los datos son adecuados y útiles para el uso que se le pretenda

dar, de esta manera depende de un factor que de antemano no conocemos específicamente, pero que de acuerdo a la clasificación de usos y usuario se puede tratar de intuir.

- Indicador(es)

1. *¿Los datos son pertinentes para la necesidad del usuario?*

- i) Posibles respuestas

- a) *Sí*

- b) *No*

- ii) Procedimiento

- *Paso 1: Ingresar a la página de datos abiertos*

- *Paso 2: Ubicar y descargar el archivo a evaluar*

- *Paso 3: Leer sobre el contenido del archivo*

- *Paso 3: Abrir el archivo y verificar si los datos son pertinentes*

Dimensión actualizado

Los aspectos de temporalidad en el contexto de datos abiertos se evaluará con respecto a lo que indique la reglamentación o legislación vigente en cada país. O en su defecto de acuerdo al período indicado en el portal. Por último, si no se indica ni en la legislación ni en el portal, se espera que estén actualizados respecto al mundo real. Por ejemplo, los datos de matrícula de estudiantes de primaria y secundaria serán anuales, mientras la de salarios serán mensuales.

- Indicador(es)

1. *¿Los datos publicados demuestran que la entrega se realiza en la forma que indica la ley del país?*

- i) Posibles respuestas

- a) *Sí*

- b) *No*

- ii) Procedimiento

- *Paso 1: Ingresar a la página de datos abiertos*

- *Paso 2: Ubicar el archivo a evaluar*

- *Paso 3: Verificar que los datos publicados se encuentren actualizados de acuerdo a lo establecido de la ley del país*

2. *¿Los datos publicados demuestran que la entrega se realiza en la forma que indica el portal?*

- i) Posibles respuestas

- a) *Sí*

- b) *No*

- ii) Procedimiento

- *Paso 1: Ingresar a la página de datos abiertos*

- *Paso 2: Ubicar el archivo a evaluar*

- *Paso 3: Verificar que los datos publicados se encuentren actualizados de acuerdo a lo que indica el portal*
- 3. *¿Los datos publicados demuestran que la entrega se realiza conforme a como los datos cambian en el mundo real?*
 - i) Posibles respuestas
 - a) *Sí*
 - b) *No*
 - ii) Procedimiento
 - *Paso 1: Ingresar a la página de datos abiertos*
 - *Paso 2: Ubicar el archivo a evaluar*
 - *Paso 3: Verificar que los datos publicados se encuentren actualizados conforme a como los datos cambian en el mundo real*

Dimensión claridad contextual

Es necesario tener claro qué representan los datos en el mundo real, o sea, los datos representan atributos de entidades en el mundo real. El usuario debe comprender cuál(es) entidad(es) y cuáles atributos de esta(s) representan los datos y cómo se relacionan las entidades entre sí. Otro aspecto que evalúa esta dimensión es que se encuentre registrado quién, cuándo y dónde se crearon los datos, para conocer que representan los hechos.

- Indicador(es)
 1. *¿El sitio provee los datos suficientes para tener claro el contexto de los datos?*
 - i) Posibles respuestas
 - a) *Sí*
 - b) *No*
 - ii) Procedimiento
 - *Paso 1: Ingresar a la página de datos abiertos*
 - *Paso 2: Ubicar el archivo a evaluar*
 - *Paso 3: Verificar la existencia de los datos sobre el contexto de los datos*

Dimensión contactable

Es la capacidad del sistema donde residen los datos para permitir establecer comunicación con los responsables de los datos. También se puede evaluar con la capacidad del sistema para informar/presentar los datos necesarios para establecer comunicación con el responsable de los datos.

- Indicador(es)
 1. *¿El sistema ofrece la posibilidad de comunicarse con el responsable de los datos?*
 - i) Posibles respuestas
 - a) *Sí*
 - b) *No*
 - ii) Procedimiento

- *Paso 1: Ingresar a la página de datos abiertos*
 - *Paso 2: Verificar la existencia de mecanismos que permitan establecer comunicación con el responsable de los de datos*
2. *¿El sistema brinda los datos suficientes para establecer comunicación con el responsable de los de datos?*
- i) Posibles respuestas
 - a) *Sí*
 - b) *No*
 - ii) Procedimiento
 - *Paso 1: Ingresar a la página de datos abiertos*
 - *Paso 2: Verificar la existencia de un texto que brinde los datos de contacto*

Dimensión catálogo de datos

La posibilidad de contar con un índice que facilite la ubicación de los conjuntos de datos es importante para el usuario. Más que un simple requerimiento, el hecho de que el sitio tenga un catalogo de datos es una ayuda merecedora de ser evaluada como una dimensión.

- Indicador(es)
 1. *¿Posee el sitio un catálogo de datos?*
 - i) Posibles respuestas
 - a) *Sí*
 - b) *No*
 - ii) Procedimiento
 - *Paso 1: Ingresar a la página de datos abiertos*
 - *Paso 2: Ubicar y descargar el archivo a evaluar*
 - *Paso 3: Leer sobre el contenido del archivo*
 - *Paso 4: Identificar si el sitio posee o no un catálogo de datos*

Dimensión verificable/trazabilidad

Evalúa que el conjunto de datos posea los datos suficientes para permitir realizar operaciones de auditoría, poder conocer quien lo creó, modificó, con qué objetivos. Ver las fuentes usadas para corroborar la calidad de los datos, entre otras cosas. Estos datos pueden estar en la documentación, en un archivo de metadatos, o en el mismo conjunto. Puede aparecer directamente en el sitio. El indicador se refiere al sitio, pero puede encontrarse y ser evaluada en cualquiera de las posiciones antes mencionadas.

- Indicador(es)
 1. *¿Posee el sitio los datos sobre del catálogo de datos de manera que se puedan realizar operaciones de auditoría?*
 - i) Posibles respuestas
 - a) *Sí*
 - b) *No*

ii) Procedimiento

- *Paso 1: Ingresar a la página de datos abiertos*
- *Paso 2: Ubicar y descargar el archivo a evaluar*
- *Paso 3: Leer sobre el contenido del archivo, en el sitio y en los metadatos para conocer si existen datos que permitan saber el histórico del conjunto de datos*

Dimensión volumen de datos

Esta dimensión pretende valorar el tamaño del archivo donde se encuentra almacenado el conjunto de datos. Definir un tamaño para un archivo de este tipo no es trivial, y no está dentro de nuestros objetivos. Podemos establecer como precedente la necesidad de definir un tamaño estándar para que posteriormente se analice. De forma intuitiva podemos reconocer cuatro alcances: un archivo de datos que se pueda descargar a una velocidad promedio puede tener un tamaño inferior a doscientos cincuenta megabytes, esto es fácilmente manipulable. Segundo, un archivo de hasta un gigabyte se puede manipular de manera manejable. Tercero, un archivo de datos de un gigabyte a cuatro gigabytes es un archivo difícil de manejar. Mientras que archivos superiores de 4 gigabytes, en términos de terabytes y petabytes se pueden considerar intratables. Esta clasificación es una propuesta, ya que no encontramos literatura al respecto.

- Indicador(es)

1. *¿En qué nivel se encuentra el volumen de datos del conjunto de datos?*

- i) Posibles respuestas

- a) *Fácil de manejar*
- b) *Manejable*
- c) *Difícil de manejar*
- d) *Intratable*

- ii) Procedimiento

- *Paso 1: Ingresar a la página de datos abiertos*
- *Paso 2: Ubicar y descargar el archivo a evaluar*
- *Paso 3: Revisar las propiedades del archivo para conocer su peso*

Dimensión almacenamiento/custodia

Un aspecto necesario de medir es el que se refiere a conservar de forma permanente los datos, para realizar esta medición en el caso de los datos abiertos basta con evaluar si existen en los metadatos o en el portal alguna indicación sobre la entidad que velará por la custodia permanente y definitiva de los datos.

- Indicador(es)

1. *¿Posee el conjunto de datos o el sistema el nombre de la organización responsable de almacenar o custodiar los datos de forma permanente?*

- i) Posibles respuestas

- a) *Sí*

- b) *No*
- ii) Procedimiento
 - *Paso 1: Ingresar a la página de datos abiertos*
 - *Paso 2: Ubicar el archivo a evaluar*
 - *Paso 3: Buscar un texto en el sitio web o en los metadatos que indique la organización que velará por el almacenamiento/custodia definitiva de los datos*

4.3. Dimensiones no evaluadas

La evaluación de algunas dimensiones no se puede realizar a través de la observación, pues se requiere de conocimiento específico sobre los datos a evaluar, entre ellas:

- Exactitud
- Libre de errores
- Credibilidad (Reputación)
- Libre de sesgos (Objetividad)

La *exactitud* de un conjunto de datos se puede evaluar verificando que los valores (datos) de los atributos de las entidades sean iguales a los valores de los atributos y entidades del mundo real. Si esta dimensión se cumple en un conjunto de datos se dice que dicho conjunto está *libre de errores*.

La *reputación* de una fuente de datos tiene que ver con quien entrega los datos, si es responsable, si los entrega a tiempo, si los entrega completos, entre otras cosas. Otro aspecto que tiene que ver con la fuente es que los datos estén libres de sesgos, en este sentido se analiza si los datos son neutros, por ejemplo la entrega parcial de datos puede sesgar la apreciación al interpretar los datos.

El presente trabajo evalúa dimensiones que se pueden interpretar por medio de la observación. Las dimensiones que requieren un estudio profundo de especialistas en el dominio particular de los datos no están en el alcance de este trabajo. Se mencionan algunas pero no se pretende más que indicar que son importante pero no abordables.

Capítulo 5

Resultados y análisis

En este capítulo se estudian las posibles fuentes de recopilación de datos que se utilizarán para realizar la validación de la metodología que se ha propuesto en este estudio. Una vez definidas las fuentes se realizan las evaluaciones y se analizan los resultados.

5.1. Identificación de las fuentes de datos

La exploración de los ministerios de los países en los que se centró el estudio mostró que encontrar un grupo de ministerios homólogos no es tarea fácil. En el caso de Costa Rica se considera que el sector público está compuesto por: Poderes de la República, Órganos del Poder Legislativo, Organismo Electoral, Ministerios, Órganos adscritos a Ministerios, Órganos adscritos a la Presidencia, Instituciones autónomas, Órganos adscritos a Instituciones autónomas, entre otros.

La actual (2014-2018) administración del gobierno de Costa Rica nombró una ministra del Deporte, sin embargo, ella aparece al mando del Instituto Costarricense del Deporte y la Recreación (ICODER), siendo esta una institución autónoma, no un ministerio. No se encuentra un portal del ministerio de deporte, pero la figura de la ministra se encuentra en el portal Ministerio de la presidencia como parte del gabinete y es parte de las autoridades en el portal del ICODER como ministra. En Uruguay, la entidad homóloga es la Secretaría Nacional del Deporte, este es un órgano desconcentrado dependiente de la Presidencia de la República. Mientras que Chile sí existe un Ministerio del Deporte. Otro caso es el Ministerio de Turismo. En Costa Rica existe el Instituto Costarricense del Turismo, institución autónoma que también es liderada por un ministro; en Uruguay existe como ministerio y en Chile está fusionado con el Ministerio de Economía, Fomento. Por otra parte, en Uruguay está el Ministerio de Industria, Energía y Minería, en Costa Rica el Ministerio de Ambiente y Energía, y en Chile existe tanto el Ministerio de Minería como el Ministerio de Energía. No existe homólogo para el Ministerio de Planificación Nacional y Política Económica de Costa Rica. La figura 5.1 ilustra la dificultad para encontrar ministerios similares a los que se les pueda aplicar el experimento y que sus resultados puedan ser comparados.

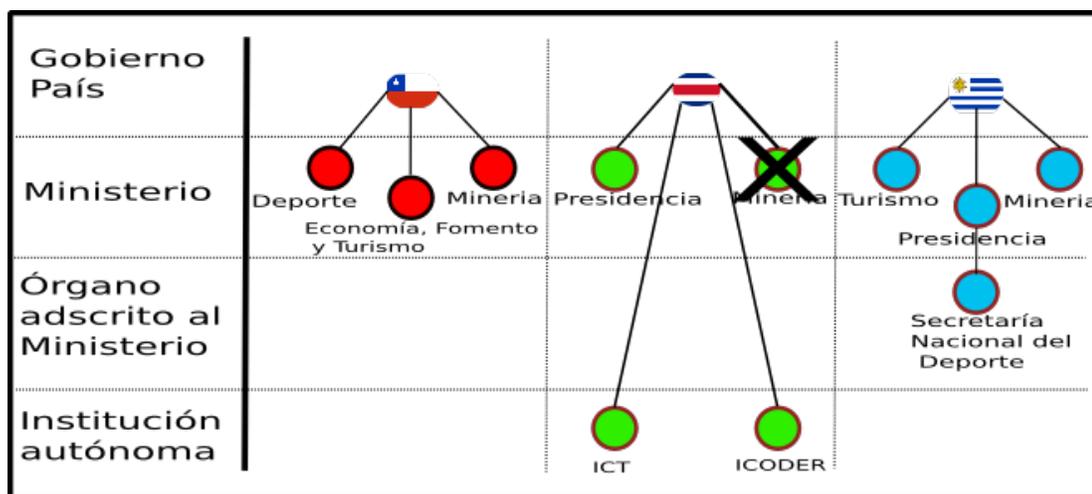


Figura 5.1: Complejidad de homologación en los Ministerios de Chile, Costa Rica y Uruguay. Siglas: Instituto Costarricense de Turismo (ICT), Instituto Costarricense del Deporte y la Recreación (ICODER)
Fuente: Creación propia

El panorama anterior nos permite reconocer que para comparar se deben identificar aquellos ministerios que realmente tengan la misma función. Para ello se consideró principalmente el nombre del ministerio, su misión y visión, además de la existencia de un portal en el que se publicaran datos sobre transparencia o datos abiertos. La observación de los sitios permitió concluir que los ministerios de Educación, Salud y Presidencia eran los que mejor cumplían esas condiciones. Para el caso de Uruguay la ley de educación en su artículo tres y cuatro del Título tercero “sistema nacional de educación pública” encargan el tema de educación tanto en manos del Ministerio de Educación y Cultura (MEC) como de la Administración Nacional de Educación Pública (ANEP), por lo que la evaluación de los sitios correspondientes a educación en dicho país se realizó en ambos portales web.

Los gobiernos de los países publican los datos abiertos a través de diferentes plataformas tecnológicas, entre las principales iniciativas de los países se destaca la creación de un portal único centralizado donde se albergan los datos por temáticas, también se pueden encontrar datos distribuidos en las secciones de transparencia de los portales de cada ministerio. Algunos datos importantes son puestos a disposición por los departamentos de estadísticas de los mismos ministerios.

Para este caso de estudio se evaluaron un total de 424 conjuntos de datos. De los 424 conjuntos de datos evaluados 109 pertenecen al gobierno de Chile, para conocer la distribución de las fuentes de Chile puede consultar la tabla 5.1. De los portales de Costa Rica se consideraron 132 conjuntos de datos y su distribución por fuente se puede ver en el cuadro 5.2. Para el caso de Uruguay se visitaron 183 conjuntos de datos y el cuadro 5.3 muestra su distribución. La tabla 5.4 permite conocer la cantidad de conjuntos de datos usados en cada gráfico. Esto ayuda a comprender el significado del porcentaje en cada gráfico.

A continuación se describen algunos de estos portales de datos abiertos en cada uno de los países donde se desarrolló este estudio.

Fuente	Evaluados	Total
Transparencia del Ministerio de Educación de Chile	21	346
Estadística del Ministerio de Educación de Chile	28	382
Transparencia del Ministerio de Salud de Chile	20	340
Estadística sobre (ENO) del Ministerio de Salud de Chile	9	240
Estadística sobre (IBS) del Ministerio de Salud de Chile	5	180
Estadística sobre (SIDA) del Ministerio de Salud de Chile	5	20
Transparencia del Ministerio de Presidencia de Chile	21	340
Total de conjuntos de datos chilenos	109	1848

Tabla 5.1: Esta tabla muestra las cantidades de conjuntos de datos evaluados y totales (la columna total es un valor aproximado) en los portales del gobierno de Chile.

Fuente: Creación propia

5.1.1. Fuentes de datos en educación

Los siguientes tres apartados describen las secciones de estadística de cada país en el tema de educación, el objetivo es que el lector conozca de donde se tomaron los datos y de qué forma se les aplicó la evaluación. Para comprender mejor la aplicación de la evaluación se entregan algunas pautas en la sección 5.1.4.

5.1.1.1. Educación en Uruguay

En el portal del Ministerio de Educación y Cultura de Uruguay el apartado de transparencia está compuesto de ocho secciones. La primera “Acceso a la información pública Ley 18.381 - Ventanilla única” consiste en dos links, uno hacia el formulario para solicitar datos e información sobre transparencia y el otro hacia una guía que indica como pedir dichos datos. La segunda sección muestra el texto de la misión, visión y lineamientos y objetivos estratégicos. La sección de “Estructura Orgánica - Decretos” también muestra textos, en ese caso como lo indica su título. La sección “Memoria Anual” muestra documentos con cuadros de datos correspondiente al período comprendido desde 2010 hasta el 2014, mientras la sección “Remuneraciones de la Unidad Ejecutora” presenta cuadros con datos sobre los salarios de los jerarcas del ministerio. La sexta sección contiene cuadros que resumen las leyes y decretos promulgadas en el 2016. Las secciones “Misiones oficiales” y “Tribunales de Provisorios” presentan archivos que contienen solo texto.

La página <http://datos.gub.uy/> del gobierno de Uruguay presenta dos link a portales de datos de educación, la del observatorio de la División de Investigación, Evaluación y Estadística de la Administración Nacional de Educación Pública (ANEP) y la del departamento de estadística del ministerio de educación y Cultura.

El portal uruguayo <https://www.anep.edu.uy/observatorio/> es un observatorio que contiene estadísticas sobre los niveles educativos de la ANEP, en este se analizan tres dimensiones: Matrícula y cobertura, Personal y Resultados educativos. El primero consiste en la estadística de los establecimientos, la matrícula y la asistencia a la educación. El segundo

Fuente	Evaluados	Total
Transparencia del Ministerio de Educación de Costa Rica	20	307
Estadística del Ministerio de Educación de Costa Rica	25	120
Estadística del Ministerio de Educación de Costa Rica (HC)	18	18
Transparencia del Ministerio de Salud de Costa Rica	25	253
Estadística sobre (ENO) del Ministerio de Salud de Costa Rica	12	37
Estadística sobre (IBS) del Ministerio de Salud de Costa Rica	6	150
Estadística sobre (SIDA) del Ministerio de Salud de Costa Rica	5	25
Transparencia del Ministerio de Presidencia de Costa Rica	21	205
Total de conjuntos de datos costarricenses	132	1115

Tabla 5.2: Esta tabla muestra las cantidades de conjuntos de datos evaluados y totales (la columna total es un valor aproximado) en los portales del gobierno de Costa Rica. HC= Hoja de cálculo

Fuente: Creación propia

contiene la disponibilidad del personal docente y la forma en que se distribuyen entre y al interior de los subsistemas. El tercero detalla los resultados del proceso educativo. El gráfico 5.10 muestra el resultado de las evaluaciones realizadas a estos conjuntos de datos.

El apartado del sitio que muestra archivos de la dimensión Matrícula y cobertura ofrece setenta y dos archivos y se encuentra dividido en seis apartados:

1. ANEP
2. Educación Inicial
3. Educación Primaria
4. Educación Secundaria
5. Educación Técnica
6. Formación Docente

El apartado denominado ANEP posee trece archivos y está seccionado en tres: matrícula, que contiene tres archivos; tasa de escolarización, que contiene un archivo y porcentaje de asistentes que posee nueve archivos. El apartado educación inicial tiene seis secciones: matrícula, alumnos por grupo, establecimientos, tasa de escolarización, tasa de asistencia y porcentaje de asistentes, con doce conjuntos. El apartado de educación primaria posee trece archivos distribuidos en las secciones: matrícula, establecimientos, alumnos por maestro, tasa de escolarización, tasa de asistencia y porcentaje de asistentes. El apartado educación secundaria cuenta con diecisiete archivos, las secciones de este apartado son las mismas seis que las del apartado educación inicial. El apartado educación técnica y el de educación primaria son similares pues ambos cuentan con trece archivos, respecto a las secciones la técnica cuenta con una menos, pues la sección alumnos por maestro está ausente. El último apartado (Formación docente) es el más pequeño, ya que está compuesto de solo tres secciones: matrícula, establecimientos y tasa de escolarización. El apartado formación docente posee únicamente

Fuente	Evaluados	Total
Transparencia del Ministerio de Educación y Cultura de Uruguay	18	60
Estadística del Ministerio de Educación de Uruguay	44	397
Estadística del Ministerio de Educación de Uruguay (PDF)	15	103
Estadística de la ANEP de Uruguay	72	134
Transparencia del Ministerio de Salud de Uruguay	2	5
Estadística sobre (ENO) del Ministerio de Salud de Uruguay	5	150
Estadística sobre (IBS) del Ministerio de Salud de Uruguay	7	50
Estadística sobre (SIDA) del Ministerio de Salud de Uruguay	5	32
Transparencia del Ministerio de Presidencia de Uruguay	15	43
Total de conjuntos de datos uruguayos	183	974

Tabla 5.3: Esta tabla muestra las cantidades de conjuntos de datos evaluados y totales (la columna total es un valor aproximado) en los portales del gobierno de Uruguay.

Fuente: Creación propia

cinco archivos. Las figuras 5.2 y 5.3 presentan en resumen los archivos que se encuentran en la dimensión Matrícula y cobertura de este sitio.

La dimensión Personal es la más pequeña en comparación con setenta y dos archivos de la ya descrita dimensión de Matrícula y cobertura, y con los cuarenta y ocho de la dimensión Resultados. La dimensión personal cuenta solo con cuatro apartados: ANEP, educación inicial, educación, primaria y educación secundaria. La dimensión Personal posee sólo catorce archivos. Como las dimensiones anteriores la dimensión de Resultados se encuentra dividida en apartados y secciones. Las figuras 5.2 y 5.3 muestran en la primera columna el nombre de las secciones y sus archivos y en las siguientes columnas los nombres de los apartados donde están presentes las secciones y los archivos.

Gráfico	Fuente	Cantidad
Figura 5.6	Transparencia del Ministerio de Educación de Chile	21
Figura 5.7	Transparencia del Ministerio de Educación de Costa Rica	20
Figura 5.8	Transparencia del Ministerio de Educación y Cultura de Uruguay	18
Figura 5.9	Estadística del Ministerio de Educación de Chile	28
Figura 5.10	Estadística de la ANEP de Uruguay	72
Figura 5.11	Estadística del Ministerio de Educación de Uruguay	44
Figura 5.12	Estadística del Ministerio de Educación de Uruguay (PDF)	15
Figura 5.13	Estadística del Ministerio de Educación de Costa Rica (HC)	18
Figura 5.14	Estadística del Ministerio de Educación de Costa Rica	25
Figura 5.15	Transparencia del Ministerio de Salud de Chile	20
Figura 5.16	Transparencia del Ministerio de Salud de Costa Rica	25
Figura 5.17	Transparencia del Ministerio de Salud de Uruguay	2
Figura 5.18	Estadística sobre (SIDA) del Ministerio de Salud de Chile	5
Figura 5.19	Estadística sobre (SIDA) del Ministerio de Salud de Costa Rica	5
Figura 5.20	Estadística sobre (SIDA) del Ministerio de Salud de Uruguay	5
Figura 5.21	Estadística sobre (IBS) del Ministerio de Salud de Chile	5
Figura 5.22	Estadística sobre (IBS) del Ministerio de Salud de Costa Rica	6
Figura 5.23	Estadística sobre (IBS) del Ministerio de Salud de Uruguay	7
Figura 5.24	Estadística sobre (ENO) del Ministerio de Salud de Chile	9
Figura 5.25	Estadística sobre (ENO) del Ministerio de Salud de Costa Rica	12
Figura 5.26	Estadística sobre (ENO) del Ministerio de Salud de Uruguay	5
Figura 5.27	Transparencia del Ministerio de Presidencia de Chile	21
Figura 5.28	Transparencia del Ministerio de Presidencia de Costa Rica	21
Figura 5.29	Transparencia del Ministerio de Presidencia de Uruguay	15

Tabla 5.4: Esta tabla muestra la cantidad de conjuntos de datos que fueron evaluados para construir cada gráfico, en el caso de estudio realizado.

Fuente: Creación propia

	Inicial	Primaria	Secundaria	Técnico	Docencia	ANEP
Matrícula						
Matrícula según forma de administración.	X	X	X	X		
Matrícula según departamento y forma de administración.	X	X	X	X	X	
Matrícula en la experiencia de séptimo, octavo y noveno rural según departamento.			X			
Matrícula según grado y forma de administración.	X	X	X			
Matrícula de educación técnica pública según departamento.				X		
Matrícula de la experiencia de séptimo, octavo y noveno rural según grado.			X			
Matrícula de quinto y sexto de educación secundaria según orientación y forma de administración.			X			
Matrícula de educación técnica pública según nivel.				X		
Matrícula de educación primaria pública según categoría centro educativo.		X				
Matrícula de formación docente pública según opción de estudios.					X	
Matrícula de la educación pública y privada (órbita ANEP) según subsistema y forma de administración.						X
Matrícula de la educación pública (ANEP) según subsistema.						X
Matrícula de la educación privada (órbita ANEP) según subsistema.						X
Alumnos por grupo						
Alumnos por grupo de educación pública según grado.	X		X			
Establecimientos						
Cantidad de establecimientos de educación según forma de administración y departamento.	X	X	X	X	X	
Alumnos por maestro						
Alumnos por maestro de educación pública según grado.		X				
Tasa de escolarización						
Tasas brutas de escolarización de educación según forma de administración.	X	X	X	X		
Tasas brutas de escolarización de los niveles educativos de la ANEP.				X	X	X

Figura 5.2: Clasificación del apartado de Matrícula y Cobertura de la ANEP. La primera columna muestra el nombre de las secciones y sus archivos y en las siguientes columnas los nombres de los apartados donde están presente la sección y los archivos, parte 2
Fuente: Creación propia

La selección de los archivos para el caso del observatorio se realizó de la siguiente manera:

Para cada apartado en el que se divide una sección se cuenta el número de archivos, Si el número de conjuntos del apartado es inferior a cuatro se evalúa un archivo, si es de cuatro a siete se evalúan dos. Si son ocho o más se evalúan tres archivos.

Por ejemplo, la sección A se divide en tres apartados A1, A2 y A3. El apartado A1 tiene diez conjuntos, como es mayor que ocho se toman tres. El apartado A2 tiene cinco conjuntos, se consideran dos. El apartado A3 tiene dos conjuntos entonces se toma uno.

El Ministerio de Educación y Cultura del gobierno de Uruguay publica a través del área de Investigación y Estadística una serie de documentos y conjunto de datos sobre el tema de educación. Los datos se encuentran organizados por año, desde 1990 hasta el 2016. Se muestra un histórico desde el año 1990 hasta el año 2008 solo en hojas de cálculo. Mientras del 2008 en adelante se entregan en dos formatos: en archivos PDF y en hojas de cálculo. Dada esta particularidad se evaluaron los archivos considerando la diferencia de formatos. Se evaluó como máximo tres conjuntos de datos por archivos tomando en cuenta la existencia de

Tasa de asistencia							
Tasas brutas y netas de asistencia a la educación según sexo. Total país	X	X	X	X			
Tasas brutas y netas de asistencia a la educación según forma de administración, región y sexo. País urbano	X	X	X	X			
Tasas brutas y netas de asistencia a la educación según sexo. Localidades pequeñas y zonas rurales	X	X	X	X			
Porcentaje de asistentes							
Porcentaje de asistentes a establecimientos de educación según forma de administración. Total país		X	X	X			
Porcentaje de asistentes a establecimientos de educación según forma de administración. País urbano		X	X	X			
Porcentaje de asistentes a establecimientos de educación según forma de administración. Localidades pequeñas y zonas rurales		X	X	X			
Porcentaje de asistentes a establecimientos de educación públicos y privados. Total país.	X						
Porcentaje de asistentes a establecimientos de educación públicos y privados. País urbano.	X						
Porcentaje de asistentes a establecimientos de educación públicos y privados. Localidades pequeñas y zonas rurales.	X						
Porcentaje de asistentes a un establecimiento educativo según edades simples. Total país.							X
Porcentaje de asistentes a un establecimiento educativo según edades simples. País urbano.							X
Porcentaje de asistentes a un establecimiento educativo según edades simples. Localidades pequeñas y zonas rurales.							X
Porcentaje de asistentes a un establecimiento educativo por quintiles de ingreso según edades simples. Total país.							X
Porcentaje de asistentes a un establecimiento educativo por quintiles de ingreso según edades simples. País urbano.							X
Porcentaje de asistentes a un establecimiento educativo por quintiles de ingreso según edades simples. Localidades pequeñas y zonas rurales.							X
Porcentaje de asistentes a un establecimiento educativo por grupos de edades. Total país.							X
Porcentaje de asistentes a un establecimiento educativo por grupos de edades. País urbano.							X
Porcentaje de asistentes a un establecimiento educativo por grupos de edades. Localidades pequeñas y zonas rurales.							X

Figura 5.3: Clasificación del apartado de Matrícula y Cobertura de la ANEP. La primera columna muestra el nombre de las secciones y sus archivos y en las siguientes columnas los nombres de los apartados donde están presente la sección y los archivos, parte 2
Fuente: Creación propia

secciones dentro de los archivos, a fin de capturar una muestra significativa. Los resultados de la evaluación realizada a la sección de transparencia del ministerio de educación, a los archivos en formato PDF y a las hojas de cálculo de la sección de estadística se pueden apreciar en los gráficos de las figuras 5.8, 5.12 y 5.11 respectivamente.

5.1.1.2. Educación en Costa Rica

En Costa Rica, el sitio de transparencia del ministerio de educación posee las siguientes secciones: Información institucional, Bienes y gastos fijos, Recursos humanos, Jerarcas y decisiones, Compras y contrataciones, Presupuestos, Planes institucionales, Informes institucionales, Servicios y trámites, Participación ciudadana. En su interior estas secciones tienen diferentes estructuras y tipo de contenidos, algunos poseen apartados con texto como el caso de “servicios y trámites”, otros se encuentran divididos en apartados y en cada apartado links

a archivos de descargas como “Compras y Contrataciones”, incluso hay uno con una sola línea que es un link a una descarga, como es el caso de “Bienes y gastos fijos”. Los resultados de la evaluación de esta sección se encuentran en el gráfico de la figura 5.7.

En lo que concierne al sitio de las estadísticas de educación en Costa Rica, se identificó el portal del Departamento de Análisis Estadístico de la Dirección de Planificación Institucional del ministerio de educación. Este posee dos secciones con datos, la primera denominada catálogo de productos donde se publican los boletines por año, desde el 2010 hasta el 2017 y en ella se encuentran algunos archivos en hojas de cálculo y en PDF, mayormente en PDF. La sección llamada auto Tabulaciones muestra una lista de links a archivos para descargar, en este caso su totalidad se encuentra en hojas de cálculo. Para observar los resultados del experimento aplicado a la sección de Catálogo de productos vea la figura 5.14, si desea observar los de auto Tabulaciones vea la figura 5.13.

5.1.1.3. Educación en Chile

El portal del ministerio de educación de Chile en su sección de transparencia se encuentra dividido en veinte secciones: Marco Normativo, Actos y Resoluciones, Estructura Orgánica, Dotación de Personal, Compras y Adquisiciones, Transferencias, Información Presupuestaria, Auditorías al Ejercicio Presupuestario, Datos de Interés Ciudadano, Otros Antecedentes, Trámites del Organismo, Participación Ciudadana, Subsidios y Beneficios, Vínculos Institucionales, Antecedentes preparatorios de normas que afecten a EMT Ley N°20.416, Derecho de Acceso a la Información Pública, Costos Directos de Reproducción, Declaración de patrimonio e intereses, Información ley N° 20.285, Ley del Lobby. Estas secciones se encuentran divididas en al menos una o más subsecciones. En estas subsecciones pueden encontrarse páginas organizadas en apartados, link a archivos o estar simplemente vacíos con un mensaje indicando que no aplica. La figura 5.6 presenta el gráfico correspondiente a los resultados del experimento realizado con los datos de esta sección

El ministerio de educación de Chile también ofrece un portal de datos abiertos con estadística de educación para investigadores, el cual se encuentra dividido en cuatro secciones: Establecimientos educacionales, Estudiantes, Docentes y asistentes de la educación, y sostenedores. Los sostenedores son los dueños de los centros de estudio. Este portal en particular se encuentra en una plataforma especializada para mostrar datos abiertos. En la figura 5.9 se muestra el gráfico con los resultados de la aplicación de la metodología propuesta en este estudio.

La participación de empresas que ofrecen plataformas tecnológicas para publicar datos abiertos se fueron evidentes en el ministerio de la presidencia de Costa Rica y en este portal de estadísticas de educación. Esto facilita que los sitios ofrezcan datos con estándares de datos abiertos y con más facilidades de presentación y descarga.

5.1.2. Fuentes de datos en Salud

La evaluación de los ministerios de salud al igual que en educación se evalúa en dos subsecciones, transparencia y de estadística. Los próximos tres apartados describen los sitios de transparencia, mientras el cuarto describe la forma en que se trató la estadística, la cual se abordó considerando el contenido semántico de los conjuntos de datos.

5.1.2.1. Salud en Uruguay

De los sitios de transparencia estudiados, el sitio del ministerio de salud del gobierno de Uruguay fue el que ofreció menos conjuntos de datos. En la sección de transparencia se encontraron algunos archivos con datos sobre presupuestos, remuneraciones e informes de labor, como eran pocos archivos se evaluaron todos, es decir, no se seleccionó una muestra. Debido a la escasez de archivos con conjuntos de datos en la sección de transparencia se exploró el resto del sitio. La exploración permitió ubicar algunos archivos relacionados con transparencia fuera de la sección de transparencia. Por ejemplo, datos sobre las autoridades, misión y visión, organigramas, etc. Los resultados del estudio aplicados a esta sección se exponen en la figura 5.17.

5.1.2.2. Salud en Costa Rica

El Ministerio de Salud de Costa Rica ofrece en su sitio de transparencia las secciones que a continuación se mencionan: Sobre la red, Recursos humanos, Jerarcas y decisiones, Compras y contrataciones, Planes institucionales, Informes institucionales, Servicios y trámites, Participación ciudadana, Financiero, Accesibilidad Web, Bienes y gastos fijos. El contenido es variado, desde páginas completamente html, documentos planos de texto, archivos de word con extensión .doc hasta hojas de cálculo. Para observar el resultado de la aplicación de la metodología desarrollada en este estudio a esta sección diríjase a la figura 5.16. Una observación digna de mencionar es la desigualdad de la estructura, es decir, la diferencia en los nombres de cada apartado y la cantidad de apartados entre las secciones de transparencia de este ministerio con el del ministerio de educación del mismo país.

5.1.2.3. Salud en Chile

El gobierno de Chile a través de su ministerio de Salud presenta una sección de transparencia dividida en: Actos y documentos publicados en Diario Oficial; Potestades y Marco Normativo; Estructura orgánica y facultades, funciones y atribuciones; Personal y remuneraciones; Adquisiciones y contrataciones; Transferencias de fondos y aportes económicos entregados; Actos y resoluciones con efectos sobre terceros (patentes, permisos, derechos, concesiones, concursos públicos y otros); Trámites ante el organismo; Subsidios y beneficios; Mecanismos de participación ciudadana; Presupuesto asignado y su ejecución; Auditorías al ejercicio presupuestario y aclaraciones; Participación en otras entidades, Antecedentes preparatorios de normas jurídicas generales que afecten a empresas de menor tamaño; Lobby y gestión de

intereses; Declaración de patrimonio e intereses Ley N°20.880; Acceso a Información Pública; Otros Antecedentes; Costos de reproducción; Dificultades Técnicas, Visualizadores & Plugins. Aunque los nombres de los apartados cambia de manera poco significativa respecto al sitio de transparencia del ministerio de educación, la observación al interior de los apartados permite corroborar que el contenido es prácticamente el mismo. Sin embargo, si se compara con el sitio de transparencia del ministerio de salud de Costa Rica o con el de Uruguay la diferencia es grande. La figura 5.15 muestra los resultados de la aplicación de la metodología propuesta a esta sección.

5.1.2.4. Sobre la estadística en Salud

Ya se ha venido comentando acerca de la disparidad de los sitios y la dificultad que esto conlleva para realizar un posterior análisis comparativo sobre el nivel de apertura, capacidad de reutilización y accesibilidad de datos abiertos. Considerando esto se decidió explorar a fondo los sitios de estadística de salud a fin de encontrar temas similares. Aprovechando las solicitudes de datos que realizan entidades internacionales como la Organización panamericana de la salud/Organización mundial de la salud se logró ubicar contenidos similares en cada país. Por ejemplo, estadísticas sobre el síndrome de inmunodeficiencia adquirida (SIDA), los indicadores básicos de salud (IBS) y las enfermedades de notificación obligatoria (ENO). Estas últimas, si bien es cierto varían de acuerdo a la incidencia o brote que haya en cada país se acerca a homogeneizar los temas en los tres países donde se aplica el estudio. De esta forma, la aplicación de la metodología se realizó en cada una de estas tres temáticas en forma separada y sus resultados se pueden ver para Chile, Costa Rica y Uruguay en forma respectiva así: SIDA en las figuras 5.18, 5.19 y 5.20. Para IBS en las figuras 5.21, 5.22 y 5.23. Para ENO en las figuras 5.24, 5.25 y 5.26.

5.1.3. Fuentes de datos para los ministerios de la presidencia

El último ministerio al que se le aplicó el experimento fue al de la presidencia, este contó con la particularidad de que una parte de los conjuntos de datos del gobierno de Costa Rica se encontraban en la plataforma de entrega de datos *Junar*, la cual también se utilizó por el ministerio de educación de Chile, en la sección de estadística. En los ministerios de la presidencia de los países estudiados solamente se trabajó en la sección de transparencia, a diferencia de los otros dos ministerios en los que se contempló las secciones de estadística.

5.1.3.1. Ministerio de la Presidencia en Chile

El Ministerio de la Presidencia de Chile contó con los mismos apartados en la sección de transparencia que el Ministerio de Salud, los sitios están adoptando un nuevo estándar aunque con variantes mínimas si se compara con el del ministerio de educación que tiene la fachada que utilizaron todos los ministerios anteriormente. Presidencia en Chile cuenta con veinte apartados: Actos y documentos publicados en Diario Oficial; Potestades y Marco Normativo; Estructura orgánica y facultades, funciones y atribuciones; Personal y remuneraciones;

Adquisiciones y contrataciones; Transferencias de fondos y aportes económicos entregados; Actos y resoluciones con efectos sobre terceros (patentes, permisos, derechos, concesiones, concursos públicos y otros); Trámites ante el organismo; Subsidios y beneficios; Mecanismos de participación ciudadana; Presupuesto asignado y su ejecución; Auditorías al ejercicio presupuestario y aclaraciones; Participación en otras entidades, Antecedentes preparatorios de normas jurídicas generales que afecten a empresas de menor tamaño; Lobby y gestión de intereses; Declaración de patrimonio e intereses Ley N°20.880; Acceso a Información Pública; Otros antecedentes; Costos de reproducción; Dificultades técnicas, Visualizadores & Plugins. Como se verá más adelante estas pequeñas diferencias entre ministerios del mismo país no son muy significativas si las comparamos con las estructuras del mismo ministerio en los otros países. Para observar el resultado de la aplicación de la metodología desarrollada en este estudio a esta sección diríjase a la figura 5.27

5.1.3.2. Ministerio de la Presidencia en Costa Rica

Mientras los sitios del ministerios de salud y educación siguen la estructura propuesta el Centro de Investigación y Capacitación en Administración Pública (CICAP) de la Universidad de Costa Rica con algunas variaciones, el ministerio de la presidencia muestra una estructura muy diferente. El menú de transparencia muestra: El primer ítem fue Transparencia institucional contiene a: Sobre Casa Presidencial, Información de puestos, Contrataciones públicas, Directrices Presidenciales, Informes de Presupuesto, Planes y Cumplimiento. Segundo ítem: Otros informes. Luego el ítem Mensajes del presidente (1 DE MAYO) contiene a: Mensaje del Presidente (2014-2015), Mensaje del Presidente (2015-2016), Mensaje del Presidente (2016-2017). Luego los siguientes ítems: Portal de datos abiertos, Informes de viajes al exterior, Decretos de convocatoria, Plan nacional de desarrollo, Cooperación bcie, Mejora regulatoria, Informe cgr sobre pagos de prohibición, Informe de viajes del presidente. Mientras la sección de transparencia del sitio posee 9 apartados: Sobre esta Página, Actualidad, Sobre casa presidencial, Información de puestos, Contrataciones públicas, Directrices presidenciales, Servicios, Informes y planes, Contacto ciudadano.

El portal de datos abiertos se encuentra desarrollado en la plataforma de Junar, esta plataforma privada es manejada por terceros y vende sus servicios tanto al gobierno de Costa Rica como al de Chile. Para observar el resultado de la aplicación de la metodología desarrollada en este estudio a esta sección diríjase a la figura 5.28

5.1.3.3. Ministerio de la Presidencia en Uruguay

De los tres ministerios visitados de la República Oriental del Uruguay, la sección de transparencia del ministerio de presidencia es el que más recursos aportó. Cuenta con veintitrés apartados: Responsable, Evolución histórica y funciones, Estructura orgánica. Organigrama, Marco jurídico aplicable al organismo, Diario de sesiones, Actas Consejo de Ministros, Anteproyectos y proyectos de ley presentados, Servicios y trámites, Autoridades, Comisiones integradas, Llamados y concursos, Remuneraciones. Escala salarial, Comisiones en el exterior, Presupuesto asignado y su ejecución, Auditorías, Licitaciones, Presupuesto asignado

por organismos internacionales, Financiamiento a partidos políticos, Calendario de reuniones, Partidos y agrupaciones políticas, Resultado de las elecciones, Términos de uso y condiciones, Expropiaciones, Caso de demanda al gobierno uruguayo. Cabe mencionar, que de estos apartados ocho tienen indicaciones de no tener datos, con el mensaje: No aplica. Para observar el resultado de la aplicación de la metodología desarrollada en este estudio a esta sección diríjase a la figura 5.29

5.1.4. Algunas aclaraciones sobre la toma de datos

Esta sección aborda las dificultades más relevantes que se encontraron previo a la realización de la evaluación, así como las decisiones consideradas para elegir los conjuntos de datos. Pretende orientar al lector en la forma en que se aplicó la metodología propuesta en este estudio considerando que los conjuntos son distintos en semántica y contenido. Por esta razón se decidió estudiar la forma y no el contenido.

5.1.4.1. Dificultad de comparación

La figura 5.4 muestra los títulos de los archivos publicados en los años 2016 y 2017 por el departamento de estadística del Ministerio de Educación. Esta imagen evidencia que el orden de las entregas es diferente, los archivos entregados anualmente no son los mismos, la semántica o el significado de los títulos no es el mismo. Por ejemplo, “el boletín CNVMTS Matrícula Inicial, Exclusión y Rendimiento” fue publicado en el 2016, pero no en 2017. El caso contrario lo podemos ver en “Especialidades en la Educación Técnica” y “Estudiantes Extranjeros en el Sistema Educativo Costarricense” entre otros.

Si se considera que los conjuntos de datos que se publican en un período son distintos a los que se publican en el siguiente período en la misma sección de transparencia del ministerio de salud de Costa Rica, entonces es difícil que tengan un contenido similar al de los ministerios de salud de los otros países. Se tomó el caso de Costa Rica para partir de alguno, pero el caso de Uruguay es similar. Si se observa la figura 5.5 y se compara con la 5.4 es evidente que se está tratando con conjuntos de datos semánticamente distintos, en jerarquías distintas, es decir, prácticamente no son comparables en contenido, pues son distintas. Anteriormente se había comentado sobre la dificultad para comparar los ministerios. La figura 5.1 muestra la complejidad que hay para homologar las entidades. Estos dos ejemplos no son únicos, sino más bien son una pequeña muestra del tipo de dificultades que se encontró al intentar comparar los conjuntos de datos según su contenido.

Entendida la dificultad de comparar los contenidos de los conjuntos, se optó por la estrategia de comprobar las similitudes de forma, no de contenido. Es decir, dado un título de un conjunto de datos ubicado en el ministerio A1 del país A, no se pretendió encontrar en el ministerio B1 del país B un conjunto con el mismo contenido o con el mismo título, sino más bien se evaluaron las características de forma del conjunto en el ministerio A1 y posteriormente se evaluó un conjunto similar en las características de forma (no necesariamente en el contenido) en el ministerio B1. Ahora, ¿Cuáles son esas características/dimensiones evaluadas?

Este estudio identificó veinticuatro dimensiones: completitud, procesable por máquina, primario, estandarizado, portabilidad, capacidad de relacionarse, fácil de manejar, pertinencia, actualizado, claridad contextual, indica licencia, formato abierto, documentación, descripción, metadatos, catálogo de datos, contactable, verificable/trazable, consistencia, coherencia, sostenibilidad, privacidad, seguridad, almacenamiento/custodia. Sus definiciones se encuentran en la sección 4.2. Es claro entonces que este trabajo permitió realizar una comparación de los conjuntos de datos en términos de las características antes mencionadas y que a esto se le ha nombrado comparación de forma, sin considerar el contenido.

La única excepción en la que se realizó un acercamiento a una evaluación considerando el contenido de los datos fue el trabajo realizado en las estadísticas de la salud, que consideró los temas de síndrome de inmunodeficiencia adquirida (SIDA), los indicadores básicos de salud (IBS) y las enfermedades de notificación obligatoria (ENO) en los tres países, sin embargo, los conjuntos no eran estrictamente los mismos.

5.1.4.2. Regla general de selección

La aplicación de la metodología propuesta se realiza solo sobre conjuntos de datos de la forma descrita en la tabla 4.1, tal como se describió en la sección 4.2. En algunos casos, especialmente en los sitios de Costa Rica y Uruguay se encontraron muy pocos conjuntos de datos, por lo que al aplicar la regla general de selección se evaluaron muy pocos, cuando esto ocurrió se tomaron conjuntos de datos de forma aleatoria sin repetir hasta alcanzar al menos 20 registros o el total del universo disponible, como en el caso del sitio de transparencia del ministerio de salud de Uruguay.

La regla general de selección fue la siguiente:

Para cada apartado en el que se divide una sección se cuenta el número de archivos, si el número de conjuntos del apartado es inferior a cuatro se evalúa un archivo, si es de cuatro a siete se evalúan dos. Si son ocho o más se evalúan tres archivos. Es la misma regla usada para el Observatorio de la ANEP descrita anteriormente.

Por ejemplo, la sección A se divide en tres apartados A1, A2 y A3. El apartado A1 tiene diez conjuntos, como es mayor que ocho tomo tres. El apartado A2 tiene cinco conjuntos, se consideran dos. El apartado A3 tiene dos conjuntos entonces tomo uno.

Para el caso de los datos que se encontraron en los sitios de estadística, los cuales se encontraban en archivos de hojas de cálculo, estos estaban clasificados, típicamente primaria, secundaria, entre otras. Para cada una de esas clasificaciones se evaluó como máximo tres conjuntos de datos, a fin de capturar una muestra significativa. Se dice como máximo pues si la disponibilidad era inferior se evaluaban los que hubieran.

Para tener claridad y reproducir este experimento de forma exacta se puede acceder a la dirección de cada archivo y cada conjunto evaluado consultando el anexo correspondiente a la hoja de respuesta de las encuestas realizadas.

AÑO DE PUBLICACIÓN 2017

BOLETINES

Violencia en Centros Educativos, Curso Lectivo 2016
Estudiantes Embarazadas Menores de Edad en Centros Educativos Costarrincenses, Curso Lectivo 2016
Estudiantes Menores de Edad que son Madres en Centros Educativos Costarrincenses, Curso Lectivo 2016
Adecuaciones de Acceso y Curriculares en Educación Tradicional, Curso Lectivo 2016
Graduados como Técnico Medio, Curso Lectivo 2016
Programas de Prevención del Uso y Consumo de Drogas No Medicadas, Curso Lectivo 2016
Problemas de Salud en Estudiantes de Educación Preescolar y I y II Ciclos, 2016
Cobertura del Programa de Segundo Idioma en Educación Preescolar y I y II Ciclos, 2017
Cobertura de Asignaturas Especiales en I y II Ciclo, 2017
Matrícula Inicial en IPEC y CINDEA Curso Lectivo 2017
Estudiantes que Adelantan una o más Asignaturas III Ciclo y Diversificada Pública, 2017
Matrícula Inicial en Proyectos de Educación Abierta Curso Lectivo 2017
Especialidades en la Educación Técnica, Curso Lectivo 2017
Menores de Edad que Estudian y Trabajan o Abandonaron los Estudios por Motivos de Trabajo, 2017
Estudiantes Extrajeros en el Sistema Educativo Costarricense, 2017
Infraestructura y Servicios Básicos en Escuelas y Colegios Públicos Diurnos, 2017
Accidentes Ocurridos en Centros Educativos y Disponibilidad de Polizas, 2016-2017

AÑO DE PUBLICACIÓN 2016

BOLETINES

Estimaciones Cantonales de las Tasas Brutas de Escolaridad en el Sistema Educativo Costarricense 2001-2014
Problemas de Salud en Estudiantes de I y II Ciclos y Preescolar, 2015
Alumnos y Alumnas que estudian y trabajan, Curso lectivo 2015
Adecuaciones Curriculares en Educación Tradicional, Curso Lectivo 2016
Programas de Prevención y Uso Indebido de Drogas, Curso Lectivo 2015
Violencia en Centros Educativos, Curso Lectivo 2015
Graduados como Técnicos Medios, Curso Lectivo 2015
Matrícula Inicial en IPEC y CINDEA 2016
Estudiantes Embarazadas en Centros Educativos Costarrincenses 2015
Estudiantes que son Madres en Centros Educativos Costarrincenses 2015
Programa Aula Edad Matrícula Inicial, Exclusión y Rendimiento 2015-2016
CNVMTS Matrícula Inicial, Exclusión y Rendimiento 2015-2016
Cobertura del Programa de Segundo Idioma en Educación Preescolar y I y Ciclos 2016
Cobertura de Asignaturas Especiales en I y II Ciclo 2016
Menores de Edad que Estudian y Trabajan o Abandonaron Estudios por Trabajo 2016

Figura 5.4: Publicaciones del 2016 y 2017, sección de estadística del ministerio de educación de Costa Rica

Fuente: http://www.mep.go.cr/indicadores_edu/indice_boletines.html

5.2. Resultados y análisis

De los ministerios de educación se evaluó tanto la sección de transparencia como la sección de estadística. A continuación se muestran los gráficos de los ministerio de educación de cada país, primero los de transparencia y posteriormente los de estadística. El gráfico de la figura 5.6 se corresponde con la descripción de la sección 5.1.1.3, del ministerio de educación de Chile, específicamente presenta los resultados de la sección de transparencia

ESTABLECIMIENTOS EDUCACIONALES	ESTUDIANTES	DOCENTES Y ASISTENTES DE LA EDUCACIÓN	SOSTENEDORES
Establecimientos Educativos			
Esta sección contiene información completa de los distintos establecimientos del sistema educacional reconocidos oficialmente por el estado (educación parvularia, básica, media y especial).			
Directorio de establecimientos educacionales			
Resumen por establecimiento			
<ul style="list-style-type: none"> - A) Resumen de matrícula por establecimiento - B) Resumen de matrícula por unidad educativa - C) Resumen de docentes por establecimiento - D) Resumen de asistentes de la educación por establecimiento - E) Resumen de rendimiento por unidad educativa - F) Resumen de alumnos preferentes, prioritarios y beneficiarios SEP por establecimiento - G) Planes y programas de estudio 			
Establecimientos jornada escolar completa (JEC)			
Sistema Nacional Evaluación del Desempeño (SNED)			
Subvenciones a establecimientos educacionales			

Figura 5.5: Sección de estadística del ministerio de educación de Chile

Fuente: <http://datosabiertos.mineduc.cl/>

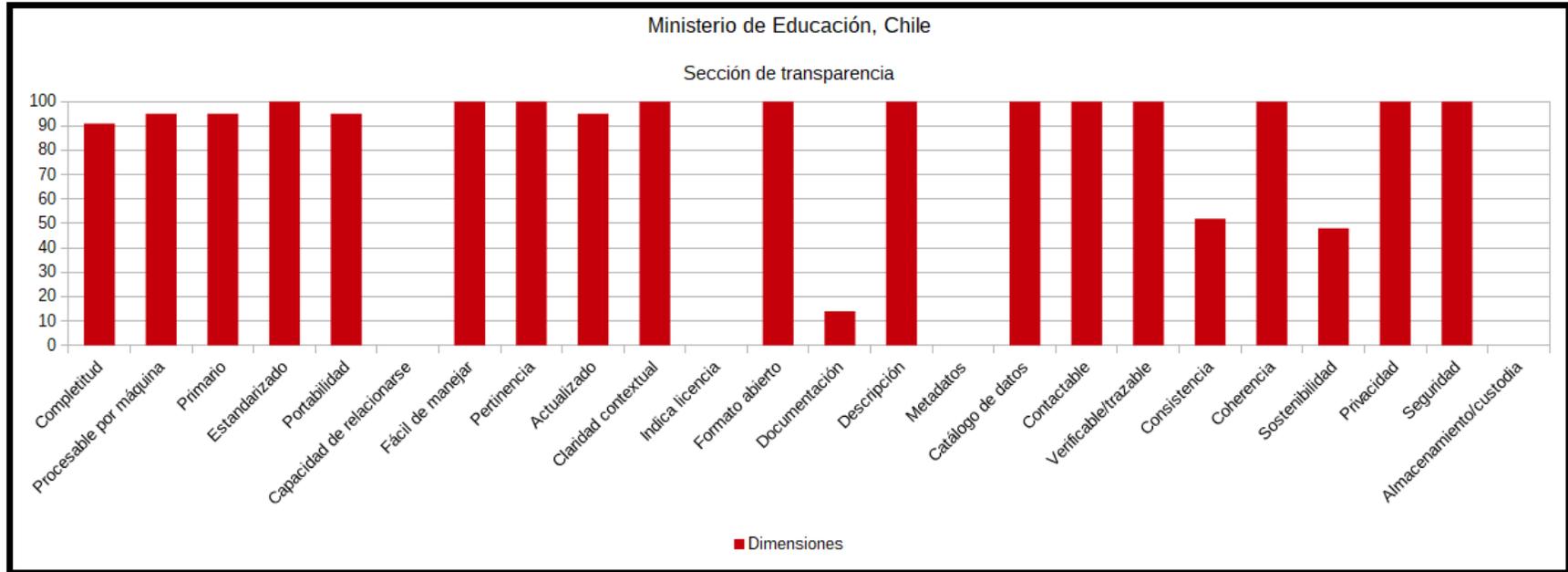


Figura 5.6: Gráfico del Ministerio de Educación de Chile, sección de transparencia
Fuente: Creación propia

En el gráfico se puede observar que hay un alto porcentaje de datos procesables por máquina, además que estos son mayormente completos, primarios y portables. Estas características permiten que un usuario técnico pueda utilizar los datos. Respecto al uso que éste pueda hacer con ellos la teoría nos mostró que la dimensión primario es importante para ofrecer la posibilidad reutilizar los datos y producir nuevos datos a partir de ellos, entonces, estos datos son útiles para producir nuevos datos. La baja consistencia y sostenibilidad muestran la poca posibilidad de utilizar los datos para utilizarlos de forma continua a través del tiempo, es decir, la teoría estudiada nos indica que estos datos no son aptos para la producción dinámica (para comprender mejor sobre producción dinámica visite la sección [3.1.1](#)).

Aunque la licencia no se indica, la posibilidad de establecer contacto es alta, por lo que la posibilidad de reutilizar no se descarta.

La sección de transparencia del ministerio de educación pública de Costa Rica fue descrita anteriormente en el apartado [5.1.1.2](#), el gráfico a continuación muestra los resultados de la aplicación de la metodología propuesta en estudio sobre la sección mencionada.

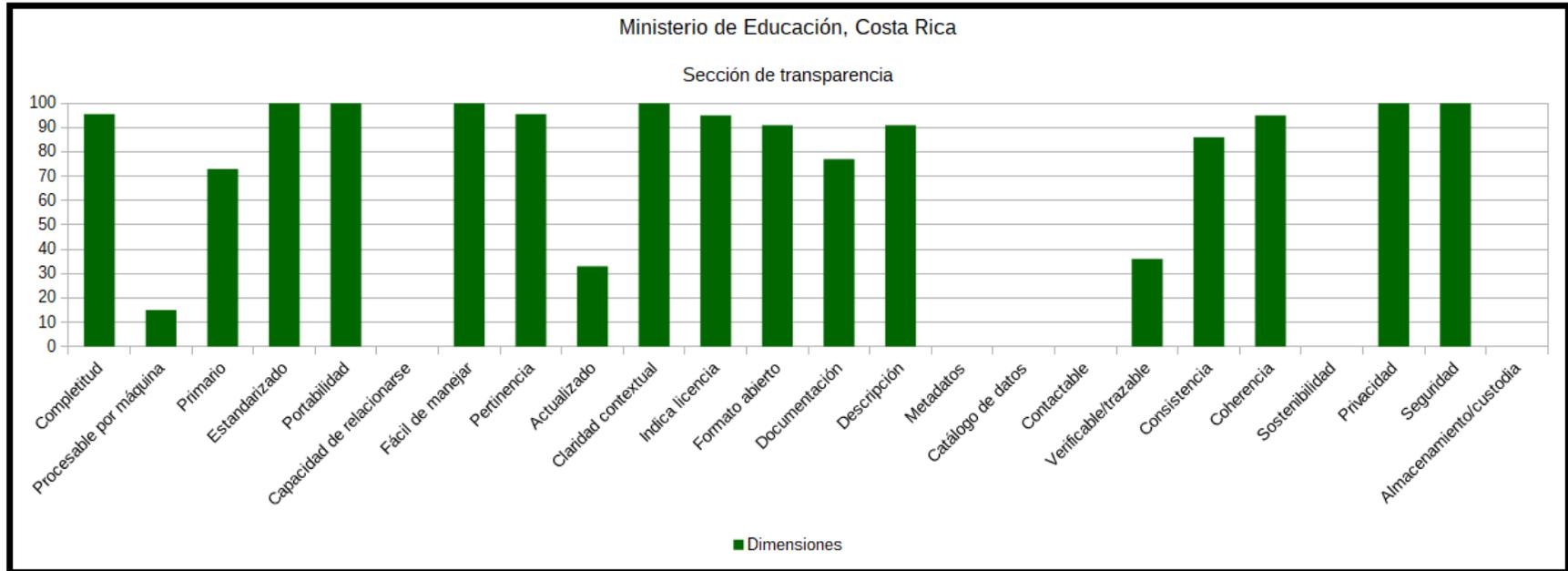


Figura 5.7: Gráfico del Ministerio de Educación de Costa Rica, sección de transparencia
Fuente: Creación propia

La figura 5.7 muestra una baja cantidad de conjuntos de datos procesables por máquinas, lo que evidencia que los datos no son útiles para usuarios técnicos, por lo que la reutilización queda en manos de usuarios no técnicos. Dado que los niveles de sostenibilidad son bajos se considera que el uso será en proyectos de corto tiempo, o sea, productos que no usen los datos de constante, que es lo que se le ha denominado producción estática. La dimensión primario alcanza casi un ochenta por ciento, permitiendo la posibilidad de reusar los datos para producir nuevos datos. Otra variable que fomenta la producción en la presencia del tipo licencia.

El tercer y último apartado de transparencia de los Ministerios de Educación es el de la República de Uruguay, la descripción de este se puede encontrar en la sección 5.1.1.1. Para el análisis del mencionado apartado se muestra a continuación su gráfico:

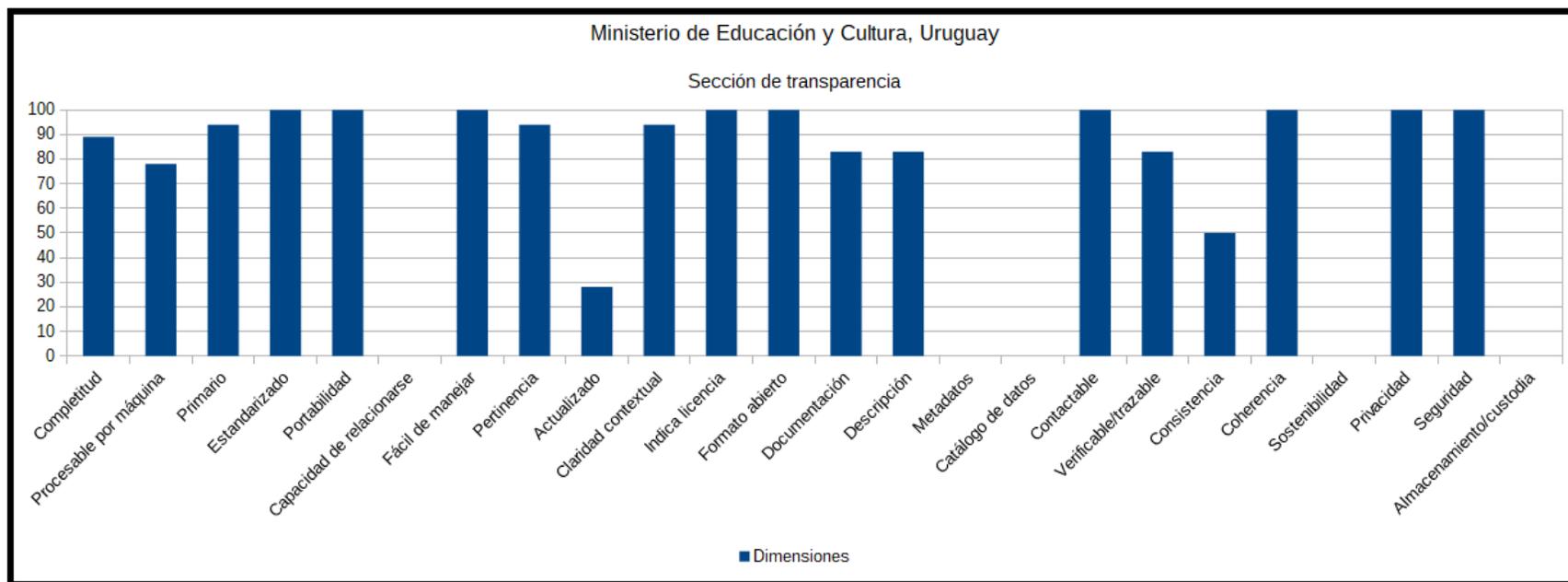


Figura 5.8: Gráfico del Ministerio de Educación de Uruguay, sección de transparencia
Fuente: Creación propia

En el gráfico de la figura 5.8 se puede observar que las dimensiones de completitud, procesable por máquina, estandarizado y portabilidad se encuentran en altos porcentajes, esto caracteriza a los sitios con facilidades para los usuarios técnicos, la cantidad de conjuntos de datos primarios también es elevado, lo que ofrece a los usuarios mayores oportunidades para producir nuevos datos a partir de estos conjuntos, una falencia de esta sección es que sus datos se encuentran poco actualizados. Por otro lado, los conjuntos indican la licencia que poseen. La sostenibilidad baja y consistencia media imposibilitan la producción dinámica y limitan solo a la producción estática.

La capacidad de relacionarse, los metadatos y el almacenamiento/custodia han sido las variables con menores porcentajes en las tres secciones analizadas, seguidos por la documentación. Por su parte, Costa Rica y Uruguay comparten la ausencia del catálogo de datos y de la sostenibilidad. Por el contrario, la variables que han obtenido los valores más elevados han sido: fácil de manejar, privacidad y seguridad.

En los siguientes párrafos se analizarán las secciones de transparencia de los ministerios de educación de los tres países estudiados. El primero es Chile, la descripción de este apartado se encuentra también en la sección 5.1.1.3, al igual que transparencia.

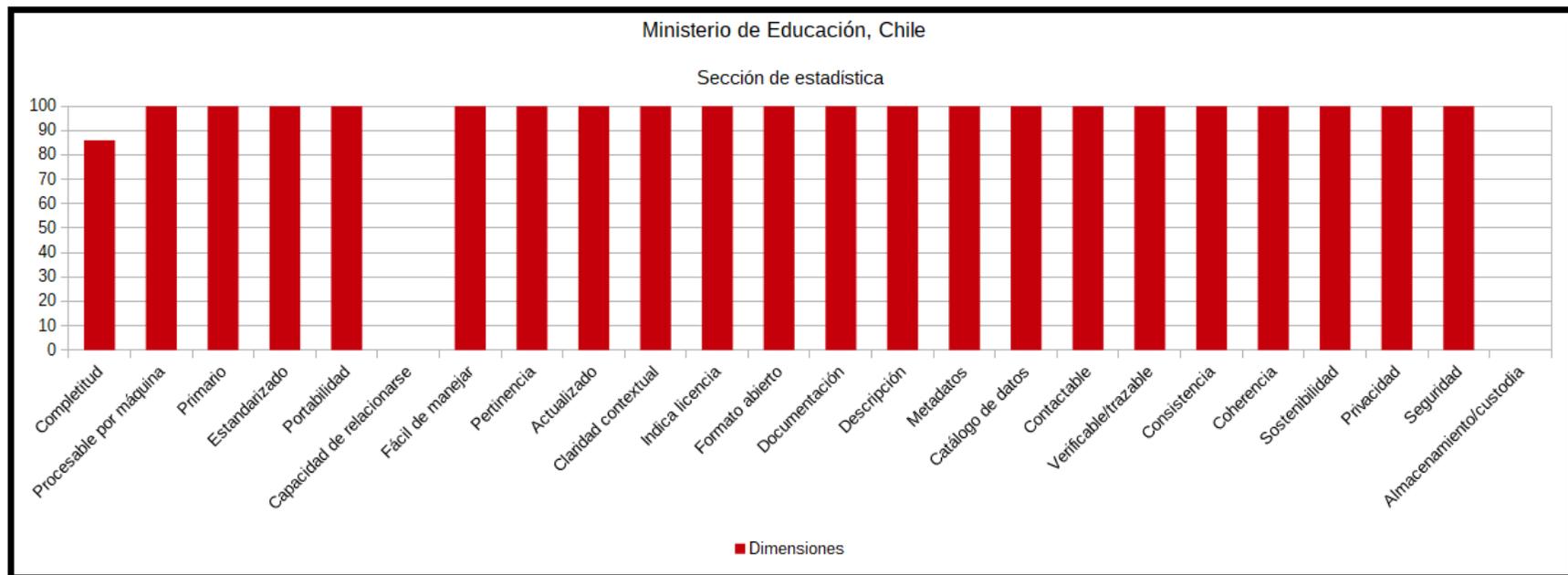


Figura 5.9: Gráfico del Ministerio de Educación de Chile, sección de estadística
Fuente: Creación propia

Al considerar la figura 5.9 se puede observar que las únicas dos dimensiones en cero son capacidad de relacionarse y almacenamiento/custodia, el resto están al cien por ciento, el alto nivel de sostenibilidad y primario permiten producir datos en forma continua, es decir con la posibilidad de que un sistema se alimenta de forma periódica a través del tiempo, esto no es en tiempo real, sino con la periodicidad que se entregan los datos, en este sentido la actualización que también es bastante alta juega un buen papel. A esta forma de producción se le ha denominado producción a largo plazo o dinámica. Como los datos son procesables por máquina, con metadatos, completos y estandarizados se facilita la manipulación para los usuarios técnicos.

De todas las secciones evaluadas esta sección es la que posee más dimensiones con valores en cien, el motivo de ello es que los conjuntos de datos se encuentran en una plataforma privada desarrollada por terceros y creada para la publicación de datos.

Como se menciona en la sección 5.1.1.1 la ley general de educación de la República Oriental del Uruguay, faculta tanto al Ministerio de educación y cultura como a la Administración Nacional de Educación Pública para ejercer autoridad sobre los temas de educación, esta última entidad ofrece un observatorio de educación, es por ello que los siguientes dos gráficos ocupan el tema de estadística de Uruguay.

Dado que la sección de estadística considera una diferencia bien marcada en términos de tipos de archivos, el proceso de experimentación se dividió dos, evaluando los archivos PDF y hojas cálculo por separados.

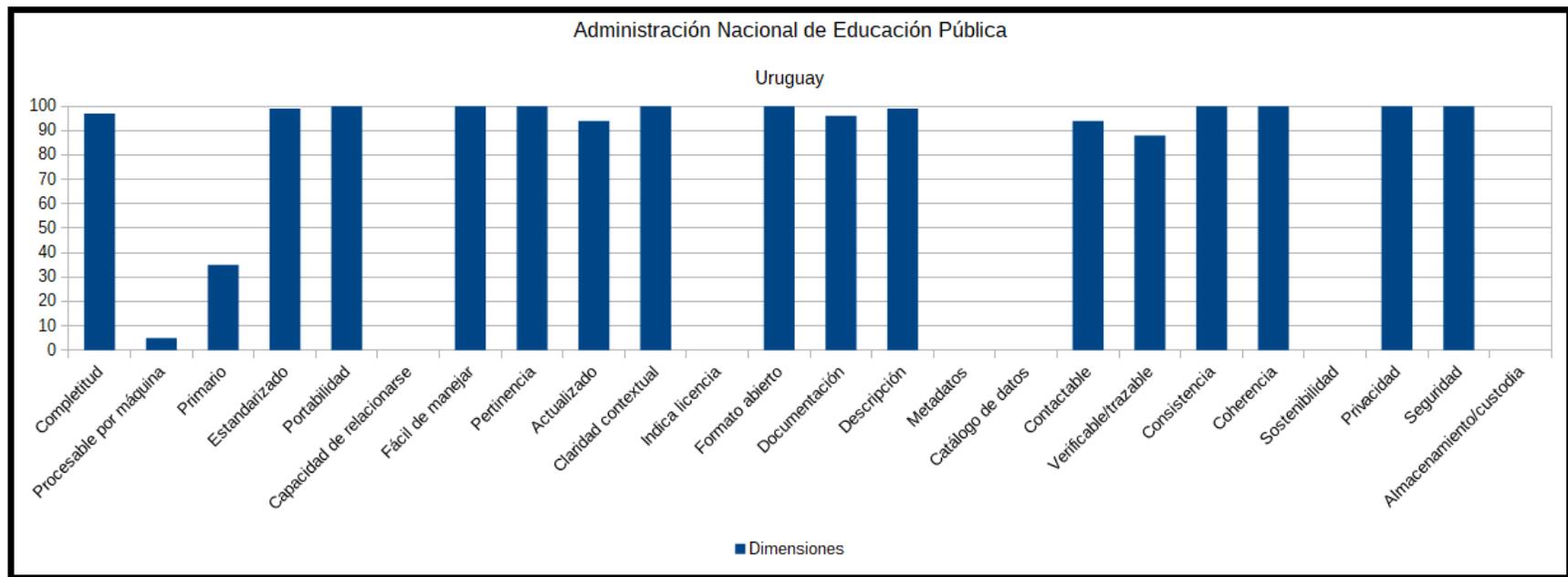


Figura 5.10: Gráfico de la Administración Nacional de Educación Pública de Uruguay
Fuente: Creación propia

La figura 5.10 nos comunica que los conjuntos de datos sometidos al experimento cuentan con un bajo porcentaje en la dimensión procesable por máquina, esto indica el usuario no técnico tiene mayores facilidades para usar estos conjuntos, también se puede apreciar que hay pocos datos primarios, por lo que no son conjuntos para producir datos a partir de ellos. Se observa que son altamente documentados y con descripciones en los conjuntos que orienta a usuarios que realizan los procesos en forma manual. No es posible realizar con ellos productos a través del tiempo, pues no son sostenibles.

Se puede observar el contraste entre las características, usos y usuarios de los conjuntos que se exponen en la figura 5.9 y la figura 5.10. Esto nos ayuda entender como el estudio de estas dimensiones nos ayudan a caracterizar/clasificar las secciones estudiadas. Por otro lado, al comparar los gráficos de las figuras 5.10 y 5.11 se concluye que se comportan de manera similar, sus principales diferencias están en que el de la ANEP tiene menos conjuntos primarios, pero más actualizados. Que tiene mayor cantidad de conjuntos en formato abierto, y que estos son más consistentes.

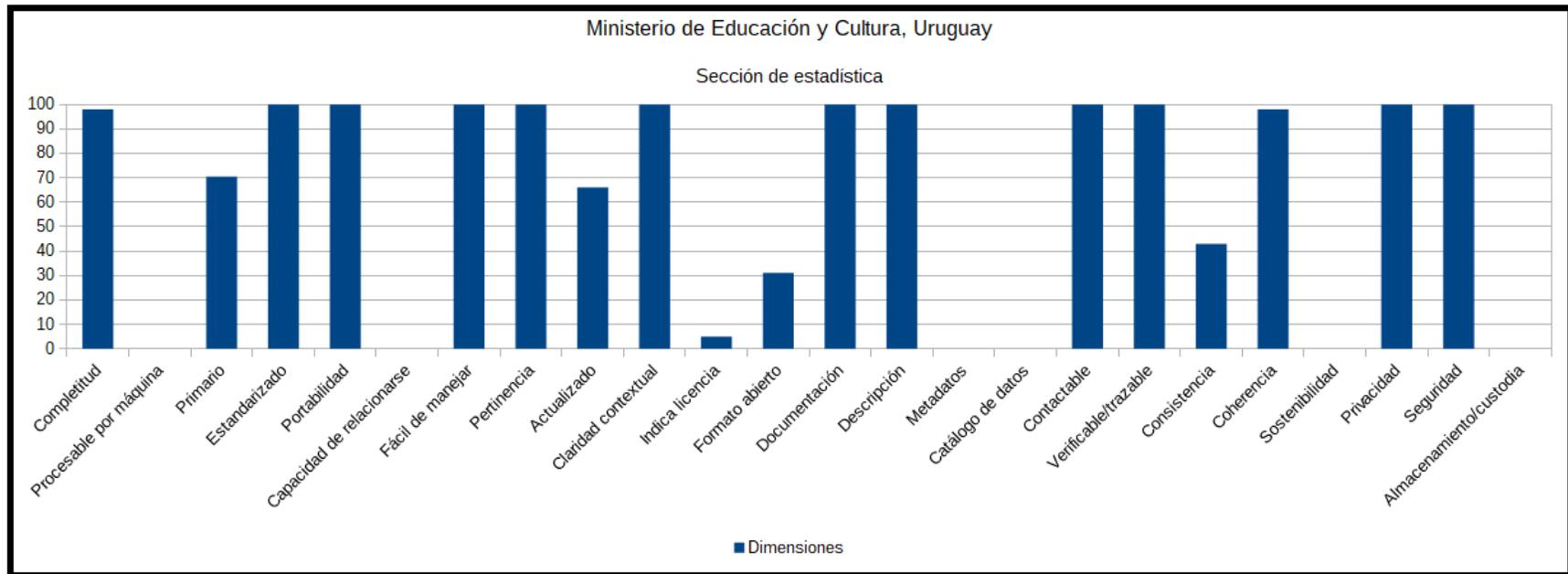


Figura 5.11: Gráfico del Ministerio de Educación de Uruguay, sección de estadística
Fuente: Creación propia

Para el caso de la figura 5.11 se puede mirar que: los metadatos, la consistencia y lo procesable por máquina alcanzan bajos niveles por lo que los conjuntos no son indicados para los usuarios técnicos. Los datos primarios son más de la mitad lo que promueve la producción de datos, además son fuertes en documentación, completitud y descripción, pero bajos en licencia y formato abierto.

La ausencia de la sostenibilidad anula la posibilidad de que los conjuntos pueda ser usados para producir datos a través del tiempo. En lo que respecta a las dimensiones de capacidad de relacionarse y de almacenamiento/custodia no han estado presente en ninguno de los conjuntos evaluados.

Vale la pena destacar que las características: privacidad, seguridad, pertinencia no son mencionadas frecuentemente en el análisis, esto debido a que otras características tienen mayor peso decisivo sobre los usos y usuarios, sin embargo si estas características no tuvieran un alto porcentaje los conjuntos no serían utilizables. En este sentido se hace necesario destacar que son dimensiones fundamentales aunque no se comenten de forma constante.

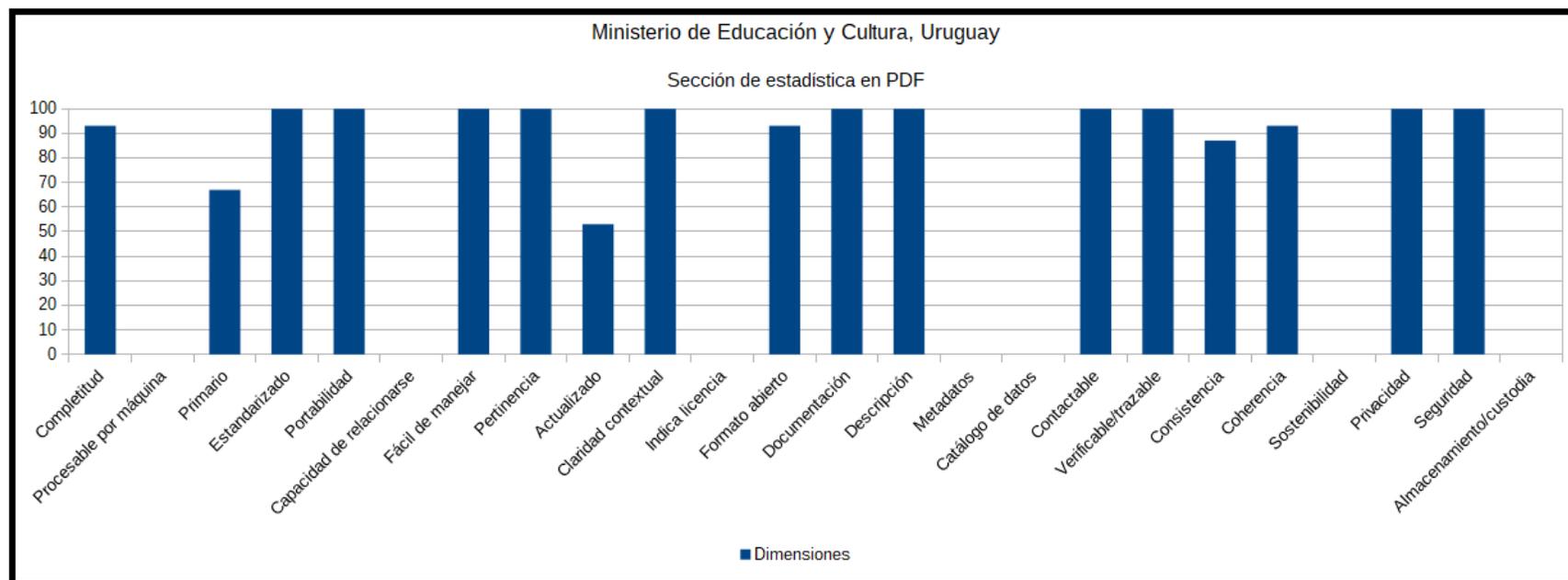


Figura 5.12: Gráfico del Ministerio de Educación de Uruguay, sección de estadística en PDF
Fuente: Creación propia

En la figura 5.12 se observa que los conjuntos no contienen metadatos ni son procesables por máquina, por lo que tiene más facilidad para trabajar con los datos los usuarios no técnicos. No hay sostenibilidad por lo que no es posible producir o consumir los datos en forma dinámica. Más de la mitad son primarios, esta dimensión es la llave para producir datos, que para este caso será de forma estática. En decremento de la producción, se puede observar datos poco actualizados y la ausencia de la indicación de la licencia.

La diferencia entre los conjuntos publicados en PDF y en hojas de cálculo es muy leve, variando solo el formato abierto y la consistencia, ambos ligeramente superiores en los conjuntos publicados en PDF.

De los ministerios de educación, solo falta analizar los gráficos de la sección de transparencia de Costa Rica. Al igual que en Uruguay el experimento se realizó en dos partes, correspondiente al tipo de archivo PDF y hoja de cálculo. La descripción de este apartado se encuentra la sección de este documento bajo la numeración 5.1.1.2. De las subsecciones de estadística la primera en evaluarse fue la de hoja de cálculos.

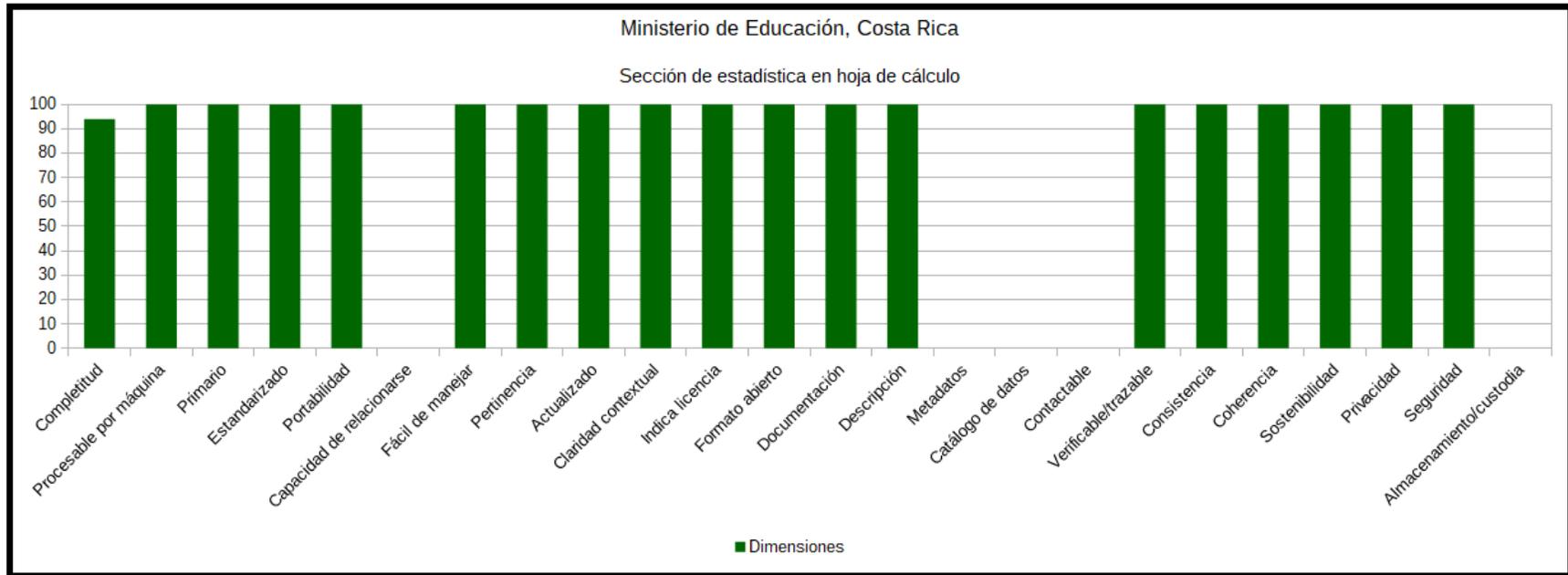


Figura 5.13: Gráfico del Ministerio de Educación de Costa Rica, sección de estadística en hoja de cálculo
Fuente: Creación propia

Considerando el gráfico de la figura 5.13 permite observar que los datos son: procesables por máquina, completos, estandarizados y portables, características de utilidad para los usuarios técnicos. Los datos primarios, actualizados, que indican la licencia, que se brinda en formato abierto permite que se utilicen para producir nuevos datos a partir de ellos. Al ser consistentes y sostenibles permite que la producción y consumo sea dinámica.

Esta sección a pesar de no estar en una plataforma desarrollada por terceros como la de Junar, es la segunda sección que tiene más dimensiones con porcentajes altos. Como ya se observó, los datos abiertos de la sección de estadística del ministerio de educación de Chile es la primera. La diferencia está en las dimensiones de metadatos y catálogo de datos, las cuales están ausentes en Costa Rica y alcanzan el máximo valor en Chile.

El último de los gráficos del tema de educación es el que muestra los resultados del experimento aplicado a los conjuntos publicados en archivos PDF.

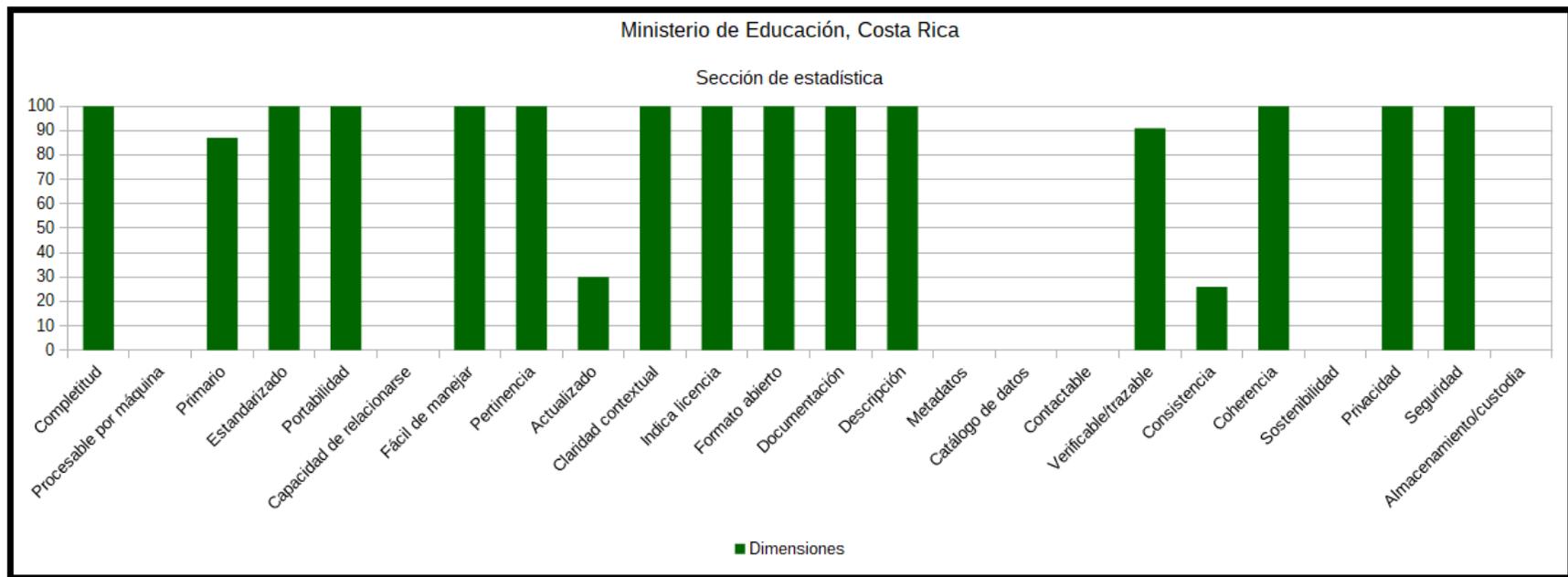


Figura 5.14: Gráfico del Ministerio de Educación de Costa Rica, sección de estadística
Fuente: Creación propia

La figura [5.14](#) muestra que los conjuntos de datos (como es de esperarse en PDF) no es procesable por máquina, lo que indica que los conjuntos están dirigidos a usuarios no técnicos. Cerca del noventa por ciento de los conjunto de datos son primarios lo que permite la producción de datos. Debido a que las dimensiones de consistencia y sostenibilidad son muy bajas no se puede producir en forma dinámica. Otro factor en contra de este tipo de producción es la carencia de datos actualizados.

La comparación entre los conjuntos de datos publicados en PDF y en hojas de cálculo permite reconocer como diferencias más significativas la ausencia de metadatos, catálogos de datos y de conjuntos procesables por máquina. Además de una leve diferencia en lo primario y una diferencia no tan leve en las dimensiones de consistencia y actualizado.

A continuación se presenta los resultados del experimento realizado a los ministerios de salud de los tres países estudiados, de forma similar a como se hizo en los ministerios de educación se trabajó en las secciones de transparencia y luego en las de estadística. El primer país en analizarse fue Chile, la descripción del sitio y sus detalles se encuentran la sección [5.1.2.3](#).

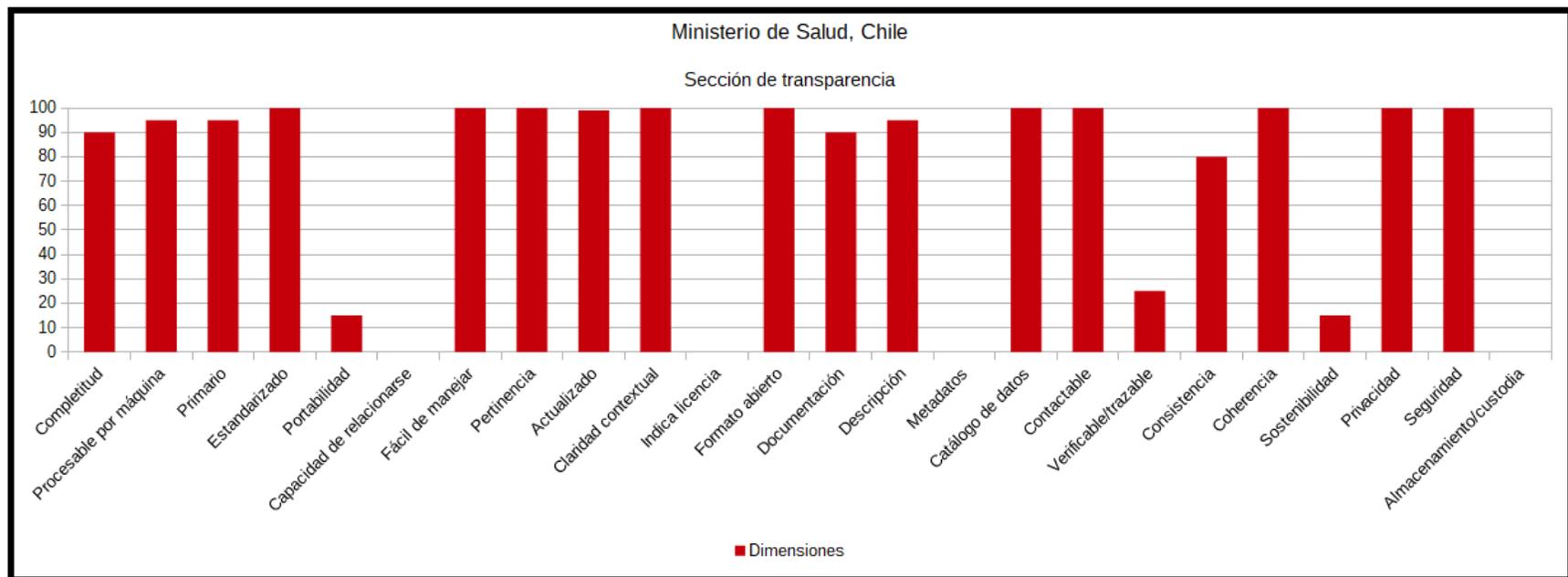


Figura 5.15: Gráfico del Ministerio de Salud de Chile, sección de transparencia.
Fuente: Creación propia

Según la figura [5.15](#) no existen conjuntos de datos procesables por máquina, no hay metadatos, no se indica la licencia y el sitio no ofrece forma ni datos para comunicarse con los responsables de los datos. Además los conjuntos están documentados, descritos, estandarizados y completos. Esto indica que los datos están dirigidos a usuarios no técnicos.

También se puede observar que el ochenta por ciento de los datos son primarios, esto abre la oportunidad de crear datos a partir de los datos que hay publicados en esta sección. Pero, la consistencia y sostenibilidad es muy baja. Por lo que la producción será solo estática.

El siguiente gráfico le ayuda comprender los resultados del apartado de transparencia del ministerio de salud de Costa Rica. La descripción del apartado mencionado se encuentra en la sección [5.1.2.2](#). Al considerar el gráfico anterior y el siguiente se percibe que los conjuntos son bastantes diferentes, sin embargo, comparten algunas características particulares como el formato abierto y la capacidad para contactar al responsables de los datos.

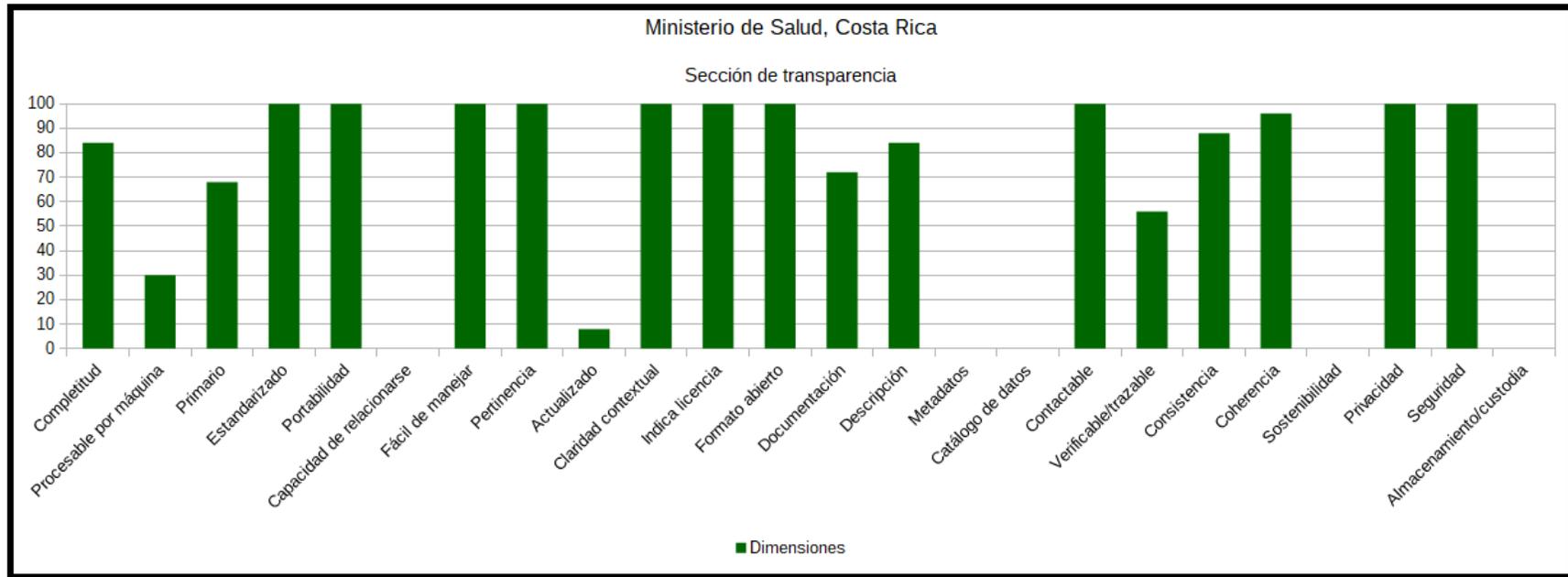


Figura 5.16: Gráfico del Ministerio de Salud de Costa Rica, sección de transparencia.
Fuente: Creación propia

Como se puede constatar en la figura [5.16](#) no hay metadatos y el porcentaje de conjuntos procesables por máquina es considerablemente bajo, esto quiere decir que la mayoría de los conjuntos son aprovechables para usuarios no técnicos. Algunas que ayudan al usuario no técnico son la documentación y descripción las cuales se encuentran cerca del ochenta por ciento.

La dimensión primaria supera el sesenta por ciento, los conjuntos indican la licencia, el formato en el que publica es abierto. Por otro lado la sostenibilidad y la cantidad de conjuntos actualizados es relativamente baja. De esta forma se espera que una buena parte de los conjuntos permita producir datos. Pero esa producción será únicamente de manera estática.

En lo que respecta a las secciones de transparencia de los ministerios de salud solo resta analizar los resultados de República Oriental del Uruguay, para conocer sobre las características y particularidades de esta sección se puede consultar la sección [5.1.2.1](#) de este documento.

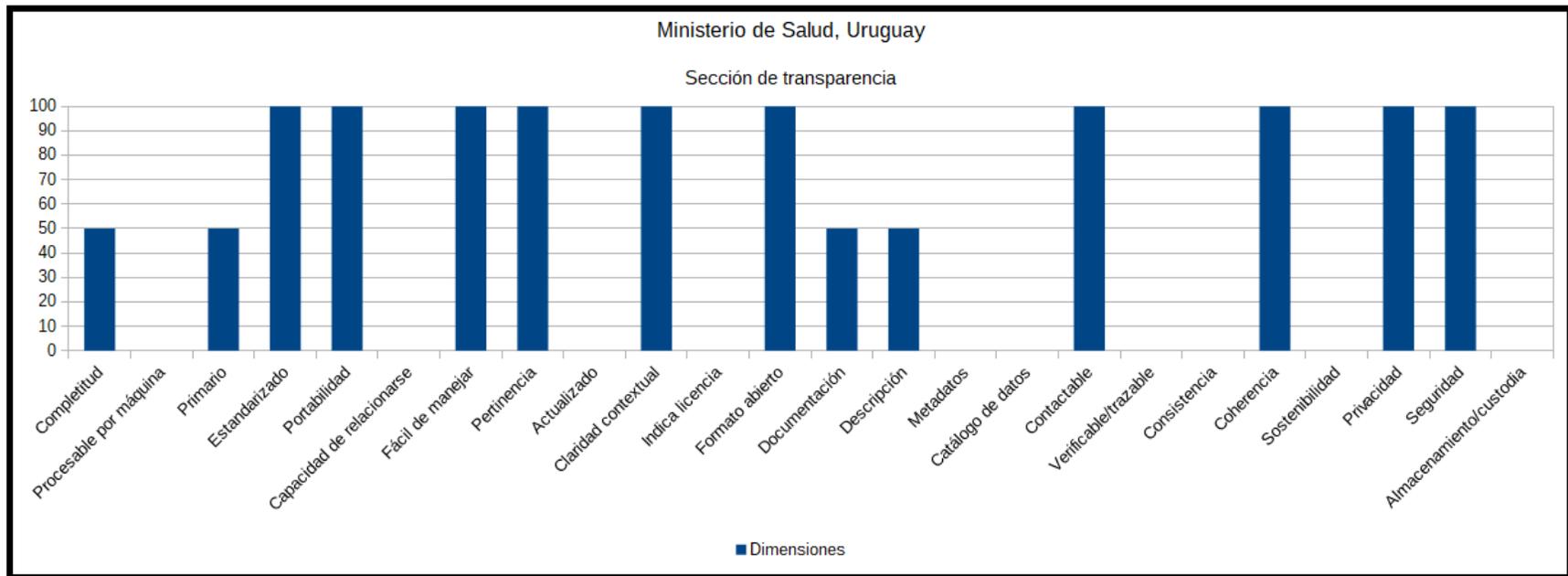


Figura 5.17: Gráfico del Ministerio de Salud de Uruguay, sección de transparencia.
Fuente: Creación propia

La figura 5.17 muestra una situación bien particular en el comportamiento de las variables que se relacionan con el tipo de usuario: no existen conjuntos que sean procesables por máquina, la completitud alcanza un cincuenta por ciento y no existe conjunto alguno que posea metadatos. Las características de documentación y descripción apenas alcanzan el cincuenta por ciento. Esto indica que los datos son insumos básicos para usuarios no técnicos que solo requieren consumir datos (no producir) y que su uso no podrá ser periódico.

La mitad de los datos son primarios, esto quiere decir que igual cantidad de conjuntos de datos se pueden utilizar para producir más datos, como para ser solamente consumidos. Los conjuntos no muestran la licencia, pero al menos permiten contactar al responsable de los datos. Ninguno de los conjuntos poseen consistencia y el nivel de sostenibilidad también es cero. De tal forma, que tanto los conjuntos de producción como los de consumo serán solamente estáticos.

El experimento realizado en los ministerios de salud de los países estudiados se realizó de forma temática. Es decir, se eligieron conjuntos de la sección de estadística de cada país que coincidieran con el mismo tema. En la sección 5.1.4.1 se expusieron las dificultades encontradas para realizar de esta forma el experimento en las demás secciones de los sitios Web. El trabajo realizado en las estadísticas de salud comprende los temas de: síndrome de inmunodeficiencia adquirida (SIDA), Indicadores básicos de salud (IBS) y Enfermedades de notificación obligatoria (ENO).

En este segmento de la presentación de resultados se exponen primero los gráficos de SIDA de cada cada país, posteriormente los de cada temática. A continuación se presentan los resultados de Chile.

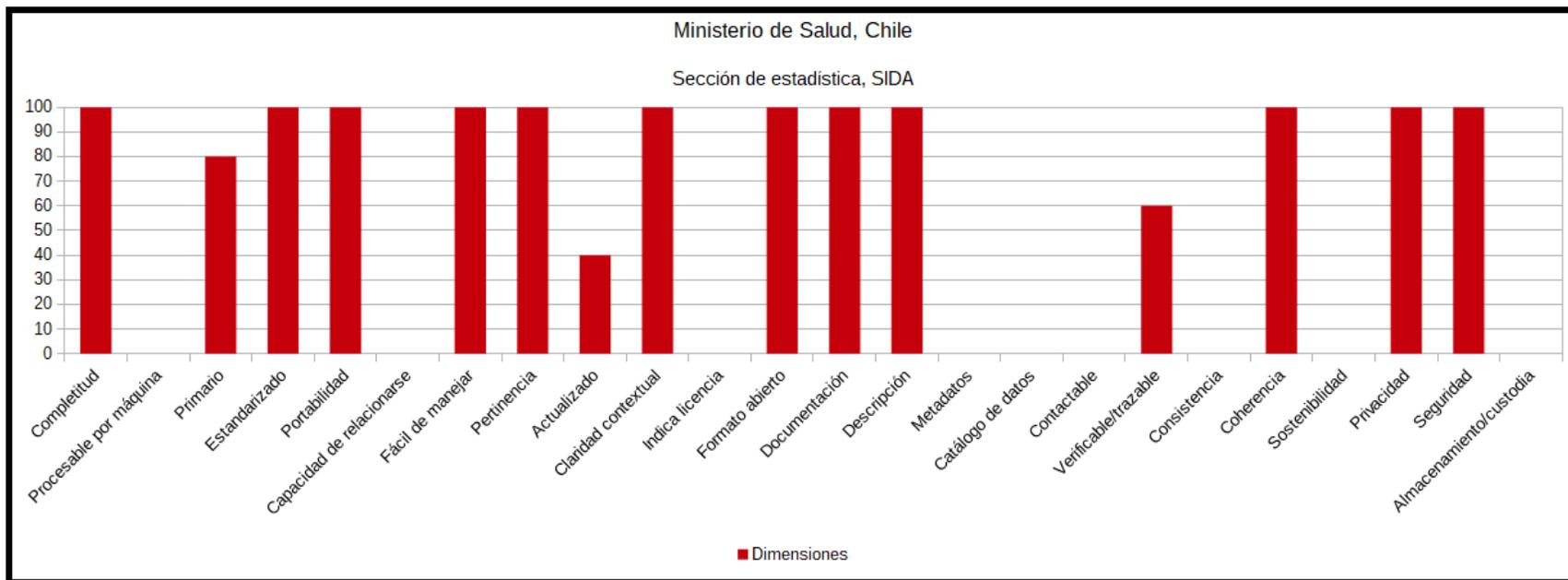


Figura 5.18: Gráfico del Ministerio de Salud de Chile, sección de estadística. Evaluación de datos sobre el síndrome de inmunodeficiencia adquirida (SIDA)

Fuente: Creación propia

Se observa en el gráfico de la figura 5.18 completitud en los conjuntos, además están estandarizados y son portables, pero no procesables por máquina, ni poseen metadatos. Sin embargo está descritos y documentados. Por lo que están publicados para usuarios no técnicos. La mayoría son primarios, mas no indican la licencia, ni existe forma de comunicarse con el responsable de los datos, y pocos conjuntos están actualizados. No existen conjuntos de datos consistentes y ninguno de sus conjuntos se publica de forma sostenible, son aptos para consumir en forma estática.

A continuación se muestra el gráfico con los resultados del experimento realizado en la sección de estadística del Ministerio de Salud de Costa Rica, correspondiente a la temática SIDA:

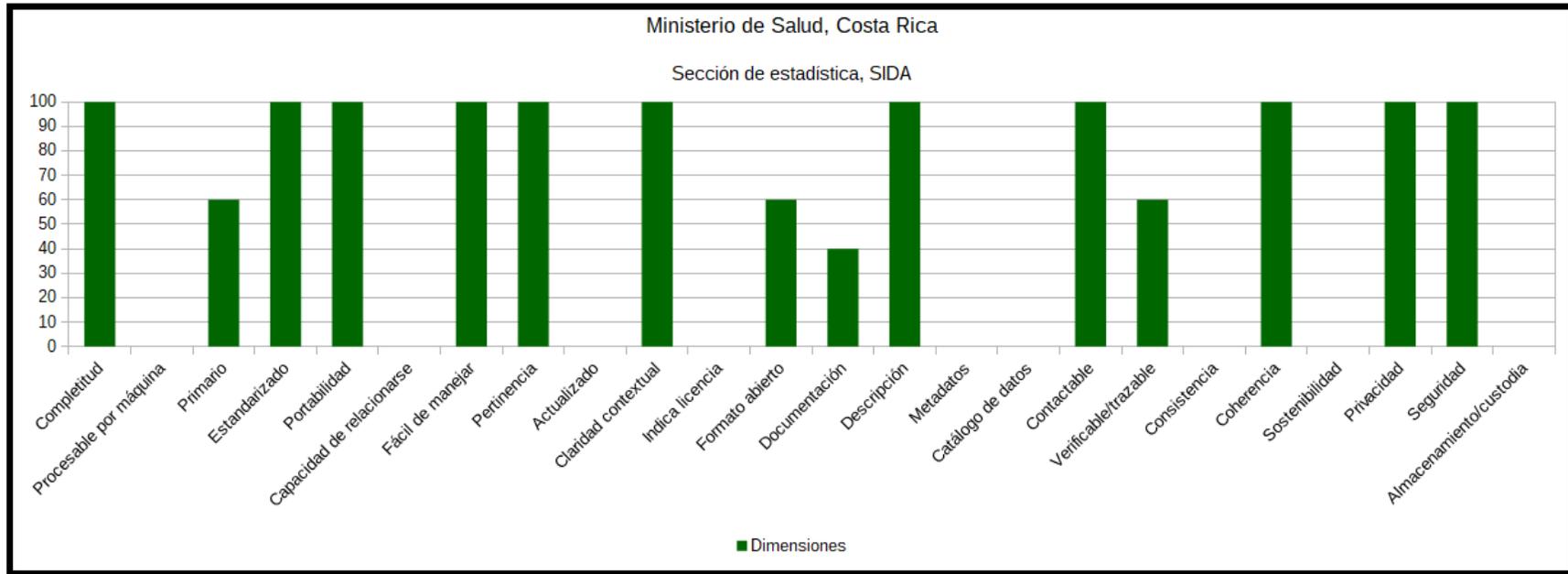


Figura 5.19: Gráfico del Ministerio de Salud de Costa Rica, sección de estadística. Evaluación de datos sobre el síndrome de inmunodeficiencia adquirida (SIDA)

Fuente: Creación propia

De la figura 5.19 se puede extraer: que existe un bajo porcentaje de conjuntos procesables por máquina, que ninguno posee metadatos, que el total de los conjuntos están completos, estandarizados, descritos, documentados y son portables. Esto permite comprender que los datos están dirigidos a usuarios no técnicos.

Por otro lado, existe la posibilidad de contactarse con el responsable de los datos. Todos los conjuntos fueron publicados en formato abierto, pero ninguno indica la licencia, tampoco existe alguno que se publique en forma sostenible, todos son inconsistentes y desactualizados. Existe un ochenta por ciento de conjuntos de datos primarios. Lo que demuestra que los conjuntos son aptos para consumir o producir de forma estática.

Para completar la triada de gráficos sobre la temática SIDA se presenta a continuación el correspondiente al gobierno de Uruguay.

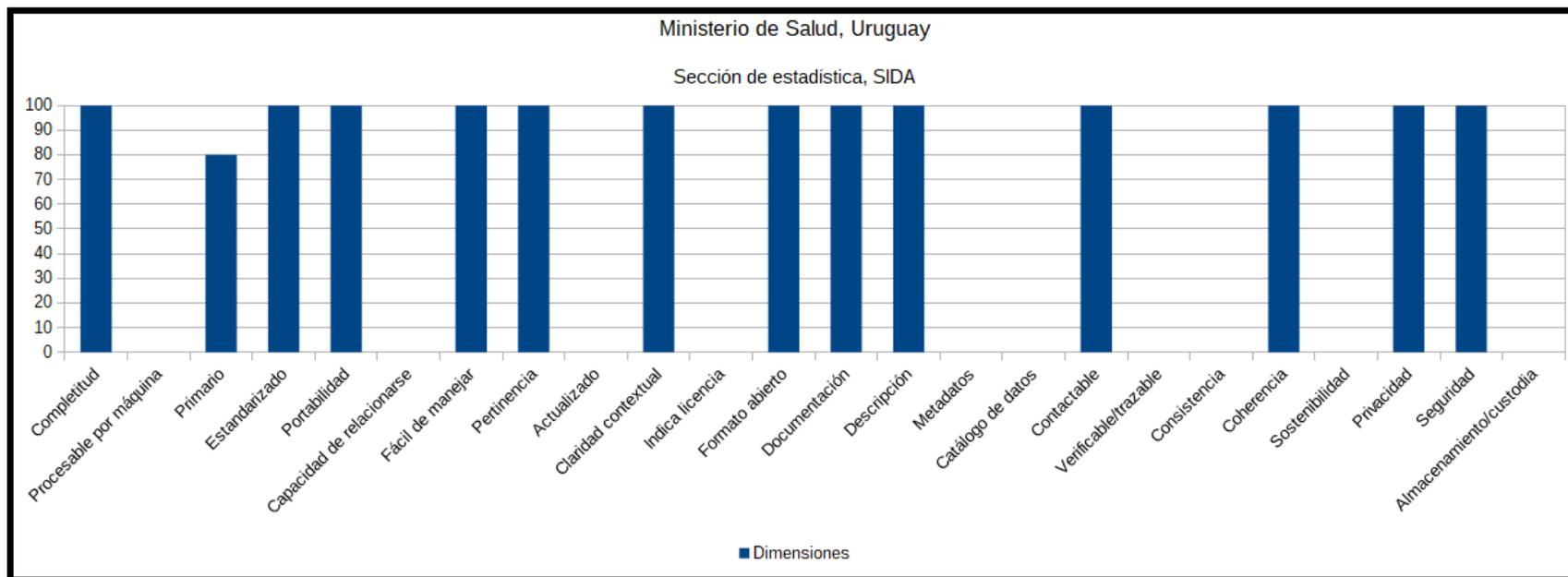


Figura 5.20: Gráfico del Ministerio de Salud de Uruguay, sección de estadística. Evaluación de datos sobre el síndrome de inmunodeficiencia adquirida (SIDA)

Fuente: Creación propia

Se puede comprender de la figura 5.20, que ninguno de los conjuntos estudiados son procesables por máquina, que ninguno tiene metadatos. Que todos los conjuntos están completos, estandarizados, descritos y documentados, y que son portables. Estos hechos comprueban que los conjuntos están dirigidos a usuarios no técnicos.

En lo que concierne a la consistencia y sostenibilidad, no existe conjunto alguno que posea dichas dimensiones. Todos los datos son publicados en formato libre y existe forma de comunicarse con los responsables de los datos, pero no se indica la licencia. Nuevamente el ochenta por ciento de los conjuntos son primarios. Los conjuntos ofrecen características tanto para ser consumidos como para producir de forma estática.

Se puede destacar de esta triada con la temática de SIDA que sus características son similares. Porque los porcentajes de conjuntos procesables por máquinas son casi nulos, salvo el caso de Uruguay que la dimensión alcanza apenas un veinte por ciento. El comportamiento de la dimensión primaria es idéntica con un ochenta por ciento en cada uno. Todos los conjuntos son portables y estandarizados con archivos de un tamaño fácil de manejar, y como es de esperarse en todos hay alta claridad contextual. Ninguno de los conjuntos de los tres países contiene metadatos, y no hay catálogo de datos en ninguno de los sitios. Los conjuntos son inconsistentes y no se garantiza sostenibilidad en ningún caso. Tampoco se indica licencia.

La segunda temática que se abordó en este estudio de las estadísticas de salud fue: Indicadores básicos de salud. A continuación se presenta el gráfico correspondiente al los resultados del experimento aplicado al gobierno de Chile.

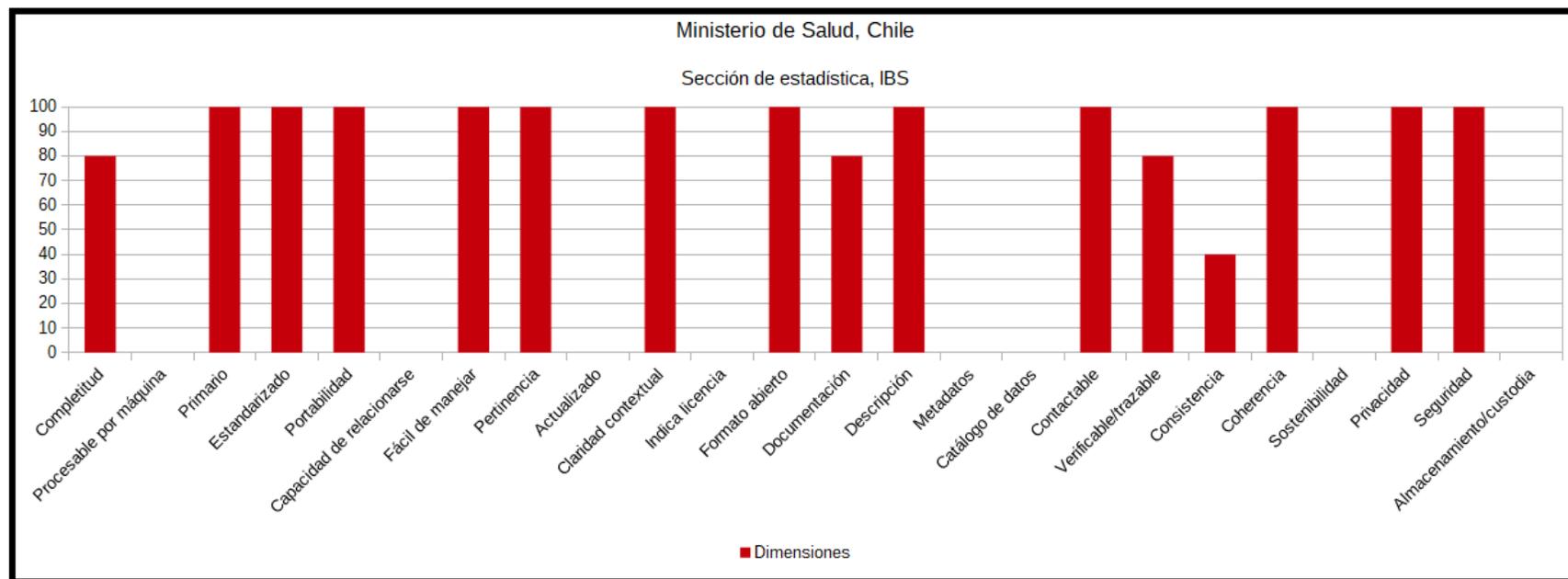


Figura 5.21: Gráfico del Ministerio de Salud de Chile, sección de estadística. Evaluación de datos correspondientes a los Indicadores básicos de salud (IBS)
Fuente: Creación propia

En el gráfico de la figura 5.21 se observa que hay un alto porcentaje de conjuntos completos, coherentes y estandarizados, que los archivos no son procesables por máquina, ni contienen metadatos. Son bastante documentados y descritos. Esto refiere a conjuntos de datos para usuarios no técnicos.

Todos los datos son primarios y con formato abierto, además es posible contactarse con el responsable de los datos, Por otro lado, no se indica licencia de los conjuntos y estos se encuentran desactualizados. Los datos no se publican de manera sostenible, por lo que el consumo o producción se limita a la forma estática.

En el siguiente gráfico se ilustra el comportamiento de las variables estudiadas en el estudio de los Indicadores básicos de salud para Costa Rica.

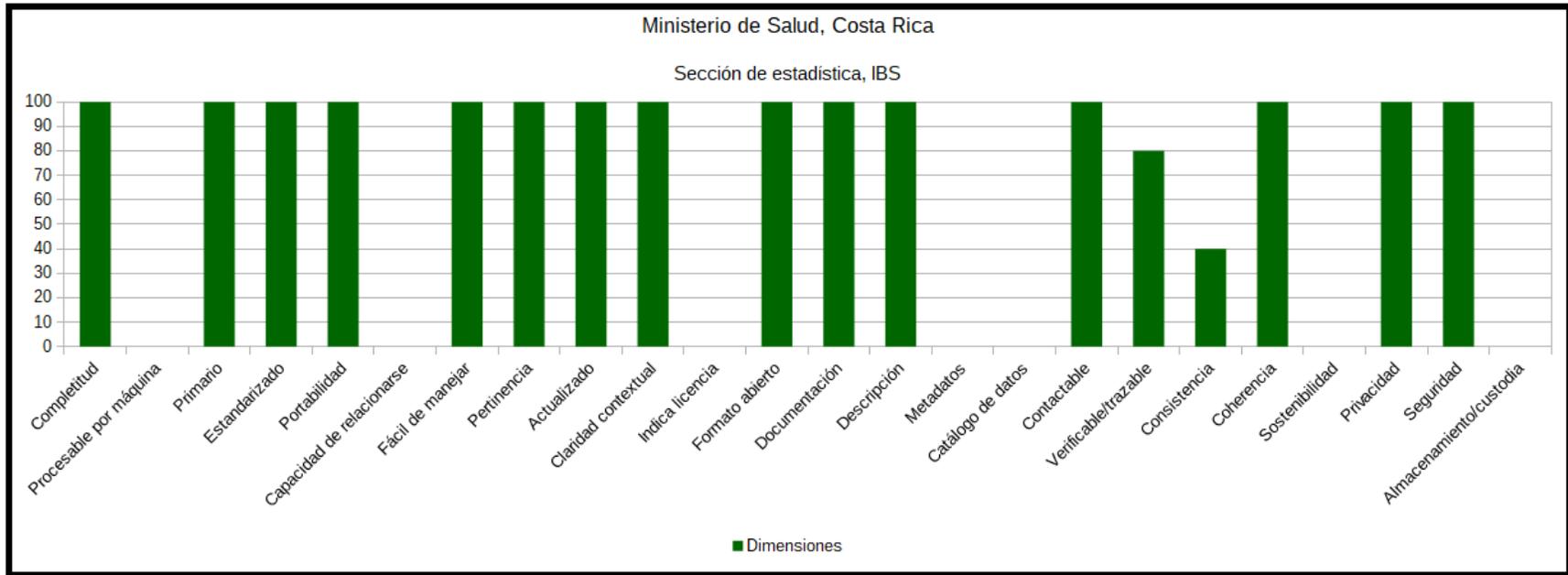


Figura 5.22: Gráfico del Ministerio de Salud de Costa Rica, sección de estadística. Evaluación de datos correspondientes a los Indicadores básicos de salud (IBS)

Fuente: Creación propia

Como se puede observar en la figura 5.22 los conjuntos están documentados, descritos, completos, estandarizados, son consistentes, son fáciles de manejar y portables, pero no contiene metadatos ni son procesables por máquina. Esto indica que los conjuntos están dirigidos a usuarios no técnicos.

No se indica licencia pero hay forma de contactarse con el responsable de los datos. Los conjuntos se publican en formato abierto, están actualizados y son completamente primarios. Lo que fomenta la reutilización de datos para producir nuevos datos a partir de ellos. Como los datos no son sostenibles a través del tiempo la producción está limitada a su forma estática.

El último de los gráficos de la temática de Indicadores básicos de salud corresponde a Uruguay y se presenta a continuación.

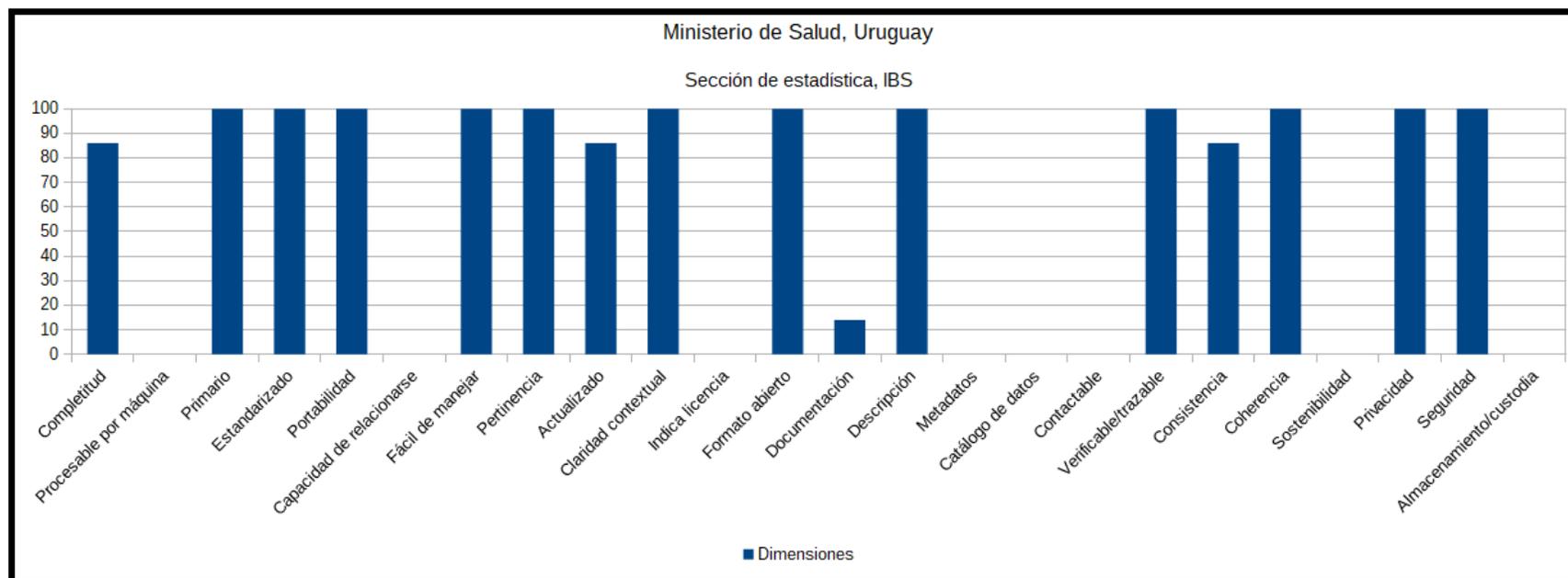


Figura 5.23: Gráfico del Ministerio de Salud de Uruguay, sección de estadística. Evaluación de datos correspondientes a los Indicadores básicos de salud (IBS)
Fuente: Creación propia

La figura 5.23 muestra que los conjuntos de son estandarizados,coherentes, portables, descritos y con un porcentaje mayor a ochenta completos. Mientras las dimensiones procesable por máquina y metadatos se encuentran en cero. Lo que demuestra que los conjuntos son aptos para usuarios no técnicos.

También se puede observar que todos los conjuntos son primarios y se han publicado en formato abierto, además son mayormente actualizados. Pero no se publican de forma sostenible a través del tiempo. Estas características nos indican que los conjuntos son útiles para producir datos a partir de ellos, pero solo de manera estática.

Una breve comparación de la triada de IBS permite reconocer que los conjuntos publicados en esta temática son cien por ciento primarios, estandarizados, coherentes, y publicados en formato libre. Ninguno es procesable por máquina ni contiene metadatos, no poseen catálogos de datos ni son sostenibles. La indicación del tipo de licencia también es nula en todos los catálogos.

El siguiente grupo de gráficos corresponden a la temática Enfermedades de notificación obligatoria (ENO), el siguiente gráfico muestra los resultados del experimento aplicado en la sección de estadística del Ministerio de Salud de Chile.

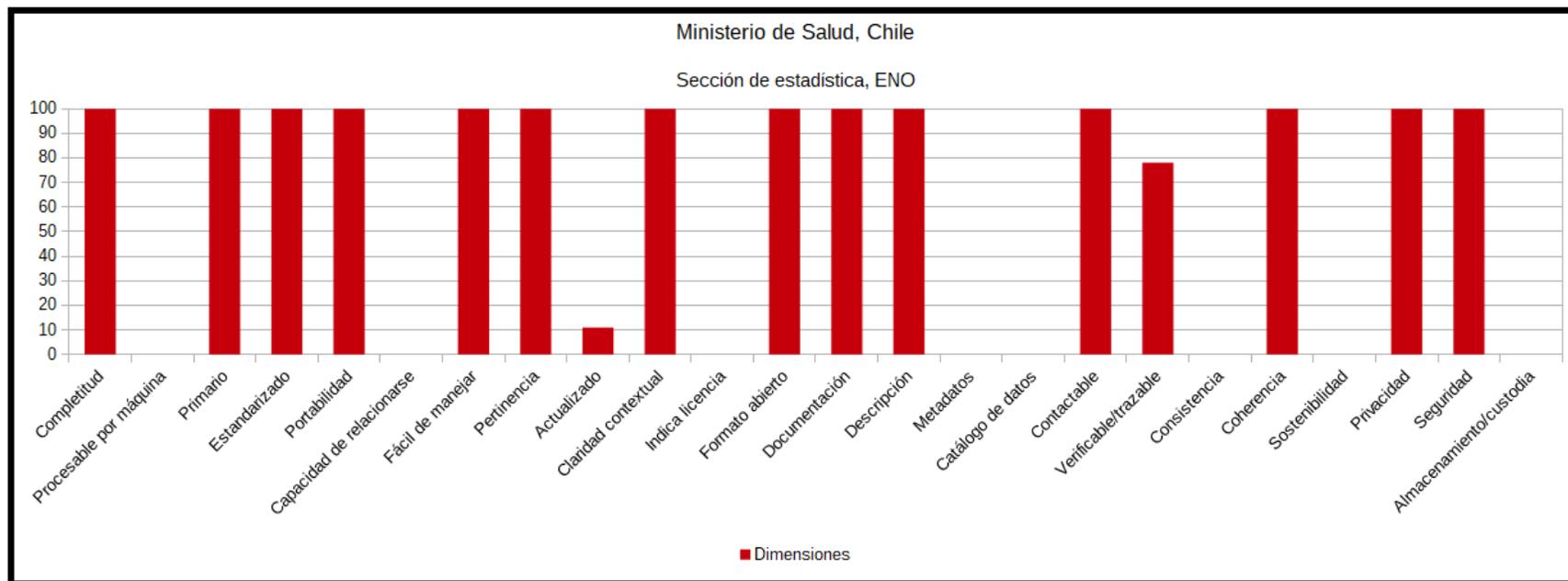


Figura 5.24: Gráfico del Ministerio de Salud de Chile, sección de estadística. Evaluación de datos correspondientes a Enfermedades de notificación obligatoria (ENO)

Fuente: Creación propia

El gráfico de la figura 5.24 permite observar que los conjuntos son completos estandarizados, coherentes, descritos, documentados, fácil de manejar. Sin embargo, no poseen metadatos ni se pueden procesar a través del computador. Estas características permiten identificar a los conjuntos de datos como apropiados para el usuario no técnico.

Se percibe también que los conjuntos son primarios, y se encuentran publicados en formato abierto. Pero son inconsistentes, desactualizados. No se indica la licencia ni son sostenibles, pero el responsable de los datos es contactable. Por consiguiente, se pueden producir datos a partir de ellos, no obstante, únicamente de forma estática.

El siguiente gráfico muestra los resultados del experimento realizado en la temática Enfermedades de notificación obligatoria en Costa Rica.

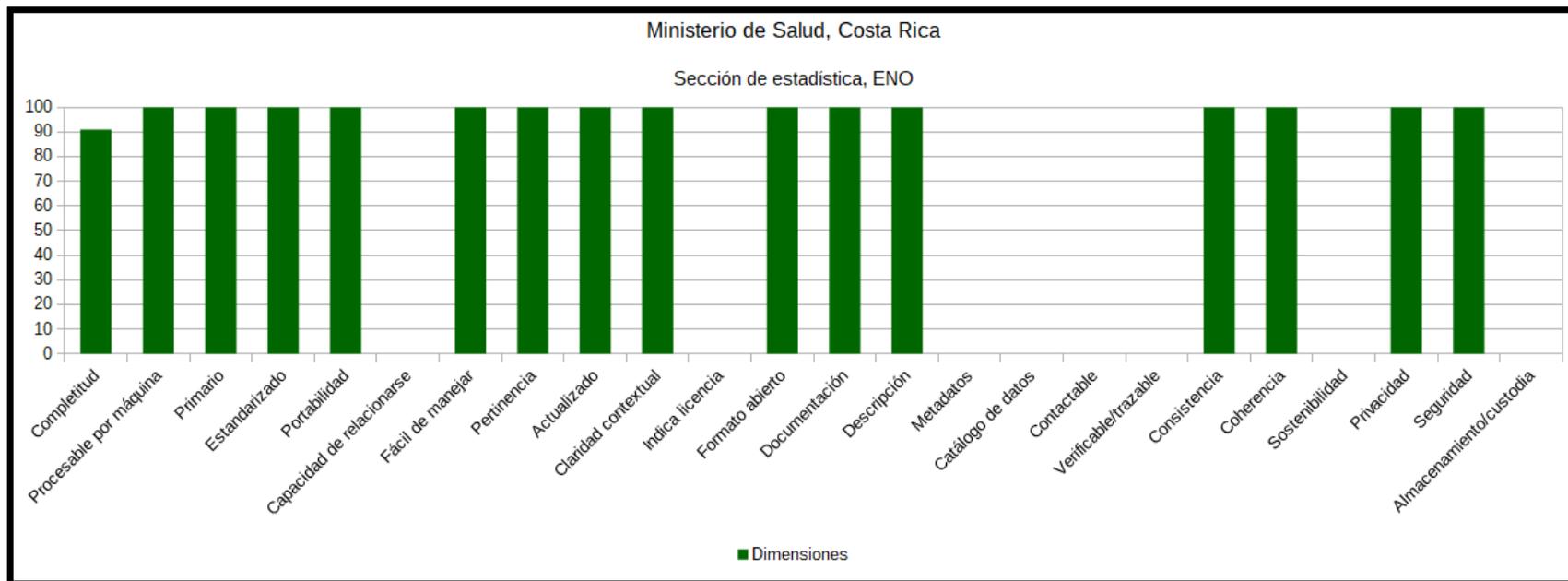


Figura 5.25: Gráfico del Ministerio de Salud de Costa Rica, sección de estadística. Evaluación de datos correspondientes a Enfermedades de notificación obligatoria (ENO)

Fuente: Creación propia

Los datos que ilustran la figura 5.25 permiten conocer que los conjuntos de datos evaluados están dirigidos a usuarios técnicos ya que: El porcentaje de completitud es cercano al noventa por ciento, los conjuntos son coherentes, estandarizados, documentados, descritos y procesables por máquina. Una desventaja para este usuario es la falta de metadatos.

El responsable los datos no es contactable y no se indica licencia. Los conjuntos están actualizados, en formato abierto y son primarios, pero no son sostenibles. Ofrecen la oportunidad de producir datos pero estáticamente.

El último de los gráficos del Ministerio de Salud y de las secciones de estadísticas corresponde al área temática de Enfermedades de notificación obligatoria del gobierno de Uruguay.

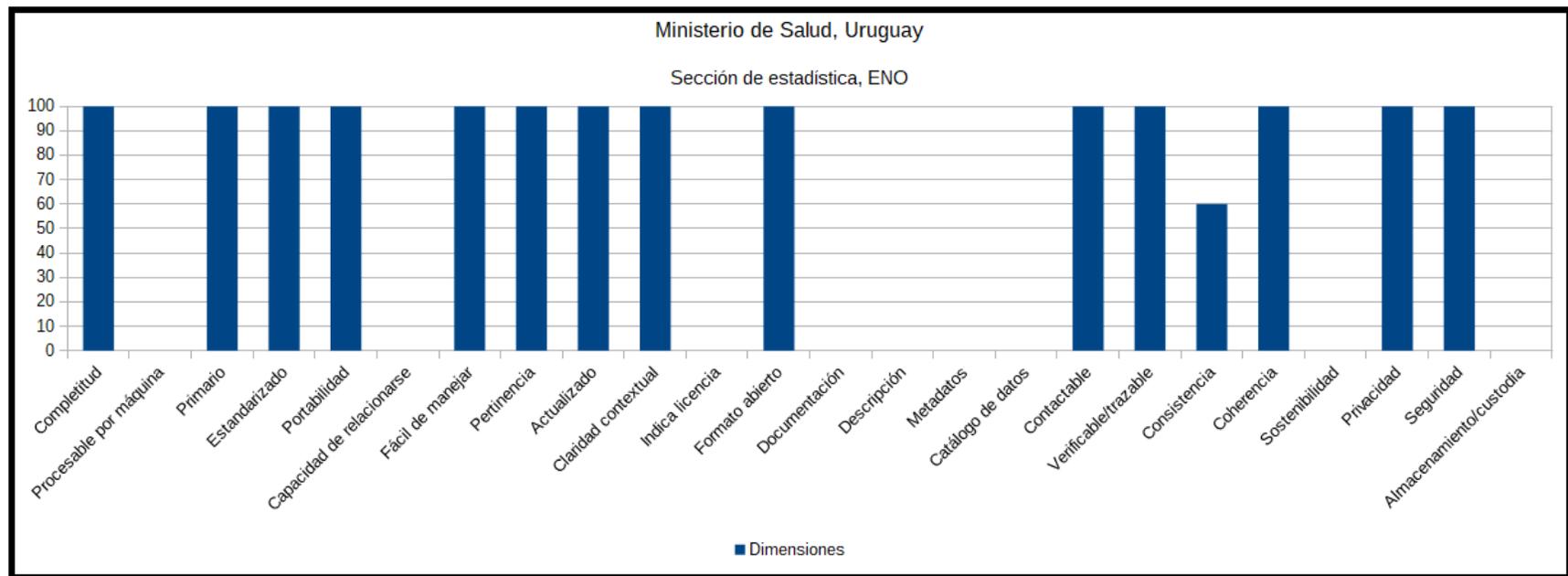


Figura 5.26: Gráfico del Ministerio de Salud de Uruguay, sección de estadística. Evaluación de datos correspondientes a Enfermedades de notificación obligatoria (ENO)

Fuente: Creación propia

El gráfico de la figura 5.26 indica que los conjuntos de datos no son procesables por máquina, pero que están completos, estandarizados, son coherentes y portables. No se encuentran documentados ni descritos y el sitio no posee catálogo de datos. Respecto a su peso son fáciles de manejar. Sus características coinciden mejor con las que encontraron útiles para los usuarios no técnicos.

Puede observarse que todos los conjuntos son primarios, que se encuentran en formato abierto y actualizados, ni el sistema ni ningún conjunto indica su licencia. Pero, es posible contactar al responsable de los datos. El cero por ciento de los conjuntos son sostenibles. De acuerdo a estas características se pueden crear datos nuevos datos y conjunto de datos, pero no de forma dinámica.

Se puede observar que esta triada de gráficos comparten algunas características, entre ellas: las tres secciones son cien por ciento: primarias, estandarizadas, coherentes, fácil de manejar, de formatos abierto. Desde otro punto de vista, ninguno indica licencia, metadatos, catálogo de datos, ni la indicación de quien custodiará los datos de forma permanente.

El tercer y último de los ministerios estudiados en las tres repúblicas fue el ministerio de la presidencia, en el cual únicamente se sometió al experimento la sección de transparencia. El gráfico que continuación se presenta corresponde a Chile y su sitio Web se encuentra descrito en la sección 5.1.3.1 de este documento.

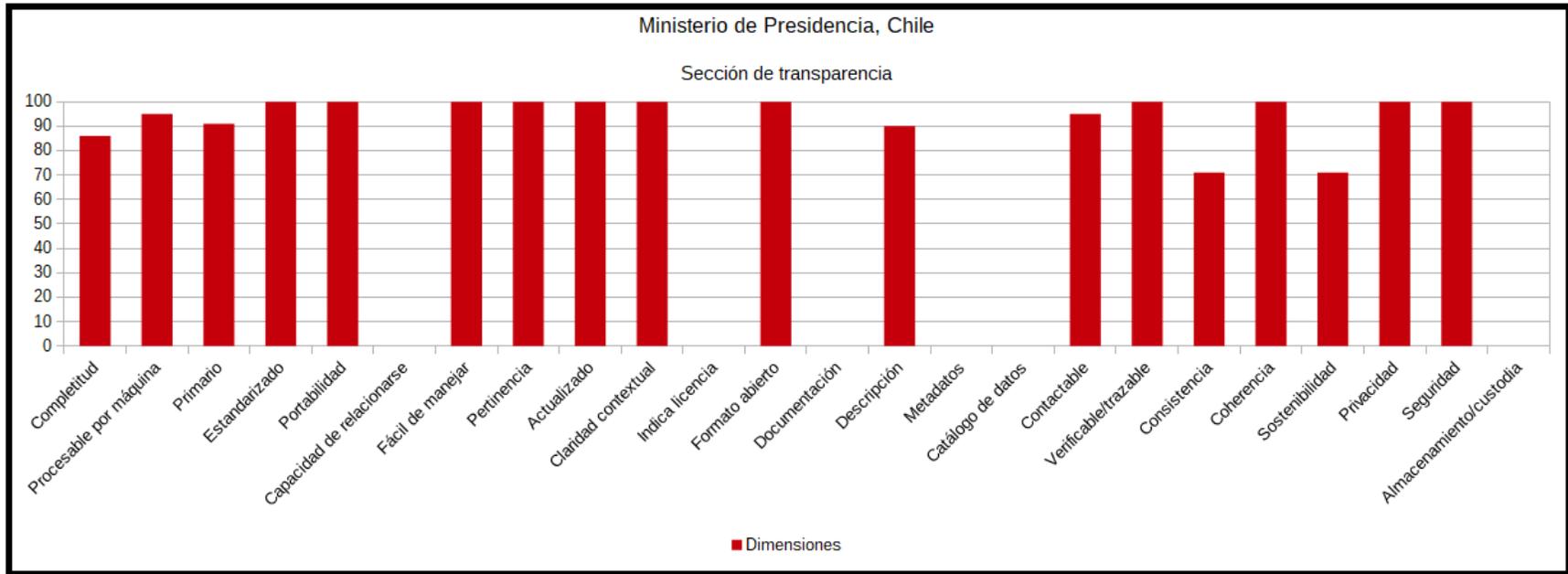


Figura 5.27: Gráfico del Ministerio de Presidencia de Chile, sección de transparencia.
Fuente: Creación propia

En la figura [5.27](#) que las dimensiones completitud, procesable por máquina, estandarizado, portable, descripción y coherencia se encuentran por encima del ochenta por ciento. Esto valores permiten identificar que esta sección está dirigida a usuarios técnicos.

Cerca del noventa por ciento de los conjuntos son primarios, todos los conjuntos están actualizados y se encuentran publicados en formato abierto. Tanto el nivel de consistencia como el nivel de sostenibilidad rondan el ochenta por ciento. Con conjuntos como éstos se pueden producir y consumir datos tanto de forma estática como dinámica.

El siguiente gráfico corresponde a los resultados del experimento realizado en el ministerio de la presidencia de Costa Rica, la descripción del sitio Web de dicho ministerio se muestra en la sección [5.1.3.2](#) del presente trabajo.

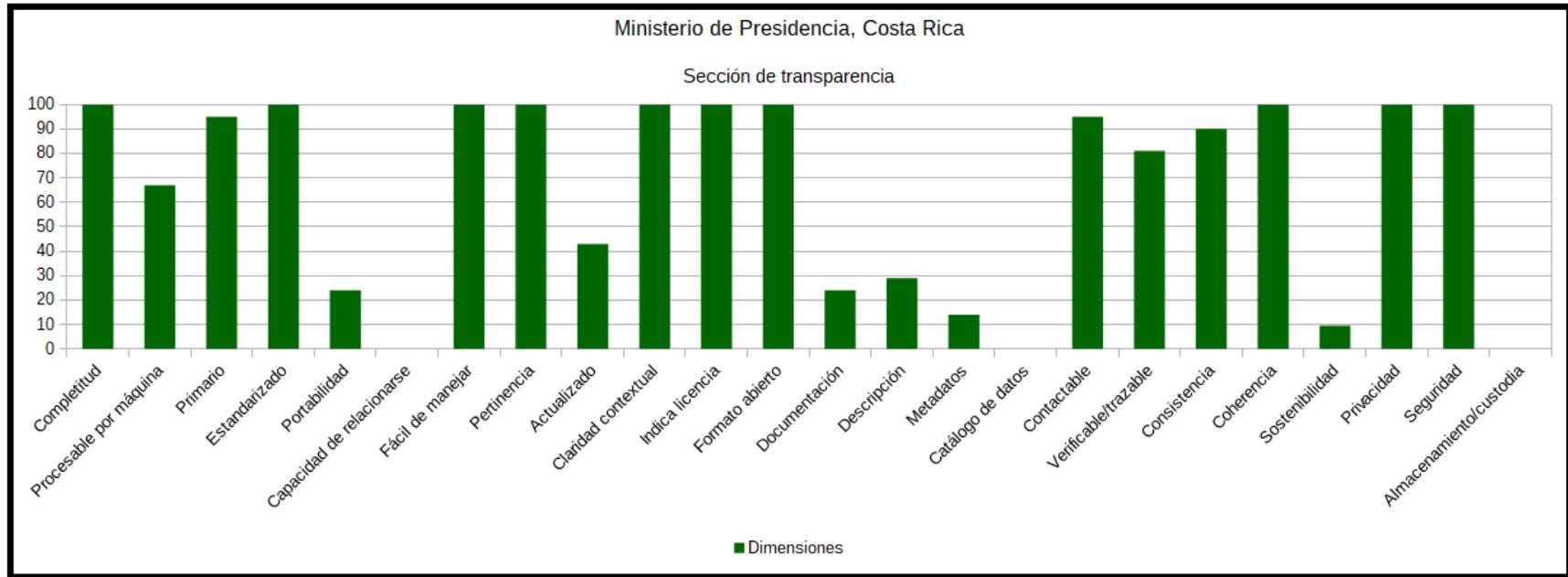


Figura 5.28: Gráfico del Ministerio de Presidencia de Costa Rica, sección de transparencia.
Fuente: Creación propia

En la figura [5.28](#) se puede apreciar que los conjuntos son pobres en documentación y descripción, existen metadatos pero también en baja cantidad. Son completos, coherentes y estandarizados. Dado que más del sesenta por ciento son procesable por maquinas se pueden considerar aptos para usuario técnicos.

Todos están publicados en formato abierto, indican licencia y el responsable de los datos es contactable, pero no son sostenibles y más de la mitad de los datos están desactualizados. Cerca del cien por ciento de los datos son primarios. Por tanto, la posibilidad de reutilizar la datos tanto para consumir como producir son altas. Sin embargo, el consumo y la producción se podrá realizar sólo de manera estática.

Finalmente, el gráfico de la sección de transparencia del Ministerio de la Presidencia de la república oriental de Uruguay cierra el análisis. En la sección [5.1.3.2](#) de este documento se describe el sitio web de dicho ministerio.

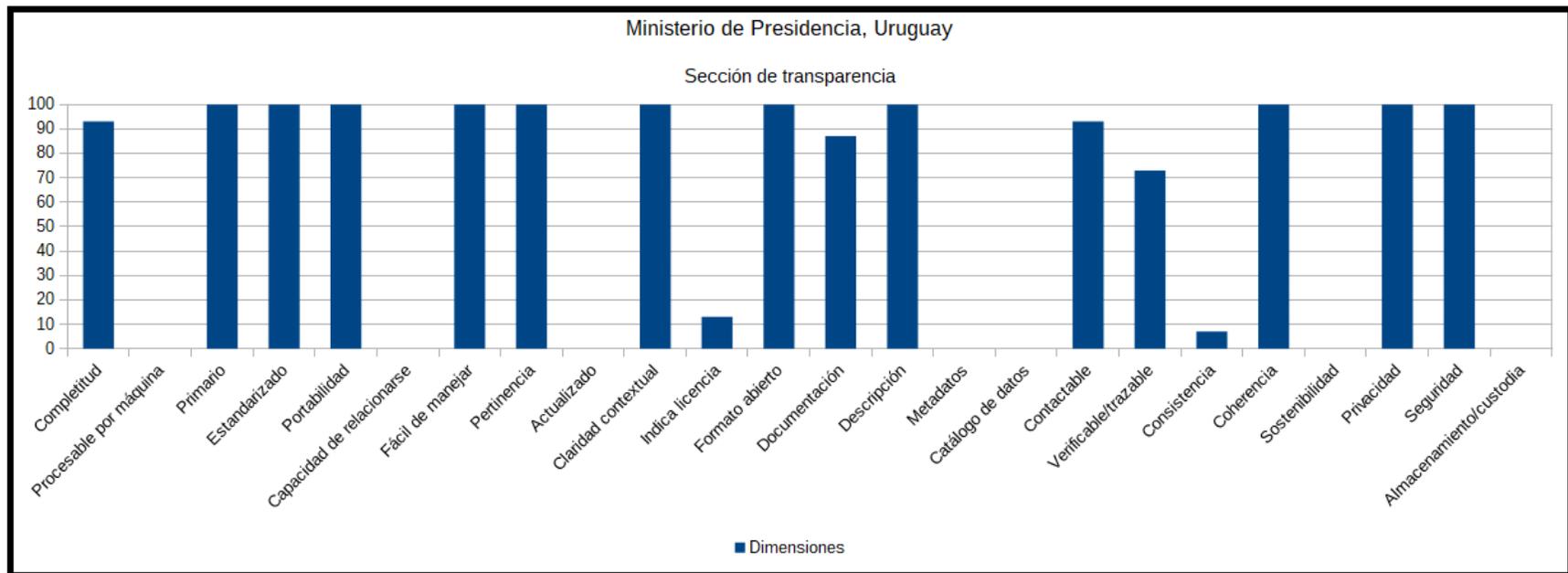


Figura 5.29: Gráfico del Ministerio de Presidencia de Uruguay, sección de transparencia.
Fuente: Creación propia

En la figura 5.29 se puede notar que las características estandarizado, portabilidad, descripción, coherencia y formato abierto alcanzan el cien por ciento, mientras los metadatos y procesable por máquina se encuentran en cero. También un alto porcentaje de conjuntos se encuentran documentados. De acuerdo con la teoría estudiada y desarrollada en este trabajo los conjuntos con estas características están dirigidos a usuarios no técnicos.

El cien por ciento de los conjuntos de datos son primarios, lo que permite que se produzcan datos a partir de dichos conjuntos. Por otro lado, la cantidad de conjuntos que indican licencia no supera el veinte por ciento y todos los datos se encuentran desactualizados. No se garantiza que los conjuntos se publiquen de manera sostenible, por lo que la forma de producción y consumo es estática.

Desde una perspectiva más amplia considerando los tres ministerios se puede notar que la estandarización, la facilidad de manejar el conjunto de acuerdo a su peso, el formato abierto y la coherencia son variables que se encuentran en todos los conjuntos evaluados. Por el contrario, las dimensiones capacidad de relacionarse, catálogo de datos y almacenamiento/custodia no se encuentran en ningún conjunto.

5.2.1. Análisis global

A continuación se realiza una comparación de manera general sobre los sitios, primeramente se comparan los ministerios entre países. Posteriormente se realiza una comparación entre países de forma general.

El primer ministerio que se compara entre países es el de educación, específicamente la sección de transparencia. Los conjuntos de datos evaluados en Chile en esta sección son altamente procesables por máquina, primarios, además cerca del 50 % son sostenibles. Por tanto, los conjuntos mostrados por Chile poseen características para usuarios técnicos productores que pueden ser usados en forma periódica. Para el caso de Costa Rica posee menos del 20 % de conjuntos de datos procesables, los datos se encuentran poco actualizados y no son sostenibles. Debido a que cerca del 80 % son primarios se pueden aprovechar para producir. Entonces los conjuntos de datos están dirigidos a usuarios no técnicos, productores y que lo utilizan de forma no periódica (única). Finalmente, el caso de Uruguay presenta mayor cantidad de datos procesables que Costa Rica, pero menos que Chile. Tiene un alto porcentaje de datos primarios, pero desactualizados al igual que Costa Rica y tampoco tiene sostenibilidad de las publicaciones en el tiempo. Considerando esto, los datos son aptos para usuarios técnicos productores y únicamente para uso único. Hay altos porcentajes de datos completos, estandarizados y descritos en los ministerios de los 3 países, sin embargo, ninguno posee metadatos.

Respecto a los sitios de transparencia de los ministerios de salud, se observa que en Chile hay un alto porcentaje de conjuntos de datos procesables, actualizados y primarios, pero un bajo nivel en la cantidad de conjuntos sostenibles. Los usuarios técnicos productores son beneficiados con estas características, mientras se destaca el uso único. En Costa Rica se repite un patrón similar al de su ministerio de educación, conjuntos de datos con porcentaje bajo en la dimensión de procesables por máquina, mayormente desactualizados y no sostenibles.

Debido a que cerca del 80 % son primarios se pueden aprovechar para producir. Las características favorecen el uso único y a los usuarios no técnicos productores. Por último, el gráfico correspondiente a Uruguay muestra que cerca del 50 % de los conjuntos son primarios. El 100 % de los conjuntos son no procesables por máquinas y están desactualizados, no indican la licencia, no son sostenibles, ni son trazables. Esto nos permite sugerir que los datos son dirigidos a usuarios no técnicos productores y pueden ser utilizados para uso único.

El último ministerio a comentar es el ministerio de la presidencia, en el caso de Chile se observa que los conjuntos son en su mayoría procesables por máquina, actualizados y primarios. La sostenibilidad supera el 60 %. Por otro lado, la documentación es nula y no se indica la licencia en ningún conjunto. Estas observaciones demuestran que los datos tienen calidades para ser utilizados por usuarios técnicos productores y con uso no periódico. En Costa Rica, los conjuntos procesables por máquina superan el 60 %, mientras los primarios son cerca del 100 %. Hay bajos porcentajes de documentación, descripción y de metadatos. Además los datos están desactualizados, la mayoría no son portables y la sostenibilidad no supera el 20 %. Estas características hacen ver que los conjuntos de datos publicados son para usuarios técnicos productores y pueden ser utilizados para uso único.

Los datos de transparencia demuestran que la capacidad de relacionarse, los metadatos, el catálogo de datos y el almacenamiento/custodia permanente son las dimensiones con porcentajes más bajos. Y que por el contrario la estandarización, la cantidad apropiada (fácil de manejar), la claridad contextual poseen altos valores.

Por otro lado, en las secciones de estadística como en el caso del Ministerio de Educación de Chile los porcentajes son más altos, ya que poseen un sistema especializado de publicación de datos llamado Junar, este también se encuentra en una parte de los conjuntos publicados por el ministerio de presidencia de Costa Rica. Algunas secciones particulares tienen también calidades altas, por ejemplo la sección de estadísticas de Ministerio de Educación de Costa Rica, esta al parecer tiene un enfoque más hacia el análisis de datos según lo indican en el propio sitio.

Finalmente, se observó que en Chile hay más conjuntos de datos procesables y más altos porcentajes de sostenibilidad, mientras que en Costa Rica y Uruguay los valores son más similares entre ellos e inferiores a Chile. Este trabajo no pretende asignar una puntuación a cada portal, si no, más bien reconocer qué características tienen los conjuntos publicados y con ello identificar para qué usos y usuarios son aptos.

5.2.2. Tablas de observaciones

Las tablas de observaciones que dan origen a los gráficos se han dejado disponible en <https://users.dcc.uchile.cl/~mvenegas/tesis>, de manera que se puede acceder a los detalles de la recolección de los datos, incluyendo las direcciones de los conjuntos consultados.

Capítulo 6

Conclusiones

Se cumplió el objetivo propuesto en la tesis, que era el desarrollo de una metodología y métricas que permitan evaluar el nivel de apertura, usabilidad y accesibilidad de los datos de gobierno, particularmente los de transparencia, centrándose en los casos de Chile, Costa Rica y Uruguay.

En la revisión del estado del arte se pudo aprovechar los diversos estudios realizados sobre *información* de transparencia, que en algunos casos refieren implícitamente a *datos* de transparencia. Aunque la literatura sobre datos abiertos es abundante, el tema de los usuarios de datos abiertos se estudia escasamente. Respecto a las *dimensiones* de datos abiertos, aunque hay literatura reciente que las aborda, ésta tampoco trata las meta-dimensiones de usuarios y usos que esta tesis incorpora.

Al estudiar los datos publicados en los sitios de gobierno, el principal desafío que surge es su presentación indiferenciada, tanto para los diferentes usos como para los diferentes usuarios posibles de esos datos. Por ejemplo, en el apartado de recursos humanos del Ministerio de Salud de Costa Rica hay datos procesables por máquina dirigidos a usuarios técnicos y datos no procesables por máquina, aptos para no técnicos. Así mismo se puede encontrar una mezcla de datos primarios para usuarios productores y datos no primarios para consumidores. Esto nos llevó a la necesidad de definir tres dimensiones principales, a saber, para quién se publica (usuarios técnicos o no técnicos); que usos se les dará (para consumo directo o para reutilización en otras aplicaciones); y por último, si el uso se realizará inmediatamente (estático) o en forma continua (dinámico).

Otra dimensión de complejidad que se encontró en los casos de estudio, fue la de *comparación* en la publicación de los datos entre los tres países, que hubo de hacerse en tres niveles: conceptual, taxonómico y de datos propiamente. A nivel conceptual, por ejemplo, en Chile existe el Ministerio del Deporte, en Costa Rica dicha entidad es una institución autónoma y en Uruguay es un órgano adscrito al Ministerio de la Presidencia. A nivel taxonómico, pues los datos no se encontraron clasificados u ordenados de forma similar. Por ejemplo, la sección de transparencia del Ministerio de Salud de Costa Rica se divide en once apartados. Mientras la de Chile en veinte. A nivel de datos, la semántica de los datos es diferente. Por ejemplo, en Chile en el apartado “Personal y remuneraciones” se publican los salarios de los funcio-

narios de forma individual con nombre y apellidos, en Costa Rica en el apartado “Recursos humanos” se presentan los datos de los salarios mínimos por profesión.

Este estudio sugiere que hay dificultades intrínsecas para estudios comparativos sistemáticos de datos abiertos de diferentes gobiernos, debido a la variedad de dimensiones, taxonomías, organización conceptual, cambios en el tiempo, que diferentes gobiernos manejan. Esta evidencia apunta a considerar los datos como parte de la vida misma de los gobiernos, y luego muchas veces irreductibles a pocas dimensiones.

Del estudio surgen diversas recomendaciones para quienes publican datos abiertos de gobierno. Primero, sería deseable que los datos en los sitios se publicaran ordenados por usuarios y por usos. Este estudio muestra que diferentes usuarios requieren diferentes tipos de formas de publicación y acceso; que los usos que se les darán pueden ser diversos y no responden a los mismos formatos, APIs, licencias, etc. Es fundamental que los publicadores de datos se esfuercen por indicar las licencias de los datos. Los publicadores deben tener en cuenta que los archivos deben poseer buenas estructuras para su óptimo aprovechamiento. Para que un archivo sea automatizable no basta solamente con un formato que facilite la automatización. Se requiere que el sitio dé las facilidades para ello, tanto en la elección de URIs, como en la APIs que exponen los datos.

Las contribuciones de esta tesis pueden resumirse como sigue:

1. Un estudio del estado del arte sobre datos abiertos y de transparencia de gobiernos, con énfasis en la región latinoamericana.
2. Un esquema de clasificación de datos abiertos, basado en la experiencia de sistemas de información y la realidad de los gobiernos, según el uso y los usuarios, que permite organizar la publicación de datos abiertos de manera más accesible.
3. Una metodología para evaluar el nivel de apertura, de usabilidad y accesibilidad de datos abiertos de gobierno que abre la posibilidad de estudios comparativos entre gobiernos.
4. Un informe preliminar sobre el nivel de apertura, usabilidad y accesibilidad de los datos abiertos en los gobiernos de Chile, Costa Rica y Uruguay.

Bibliografía

- [1] Alberto Abella, Marta Ortiz-de Urbina-Criado, and Carmen De-Pablos-Heredero. Meloda, a metric to assess open data reuse/meloda, métrica para evaluar la reutilización de datos abiertos. *El profesional de la información*, 23(6):582–588, 2014. (Citado en páginas 45 y 47.)
- [2] L. Abellán. Procesos de comunicación con perspectiva de género en el entorno de intervención, 2016. España: Nobel. (Citado en página 6.)
- [3] Russell L Ackoff. From data to wisdom. *Journal of applied systems analysis*, 16(1):3–9, 1989. (Citado en páginas 6 y 7.)
- [4] Martin Adalberto Tena-Espinoza-de los et al. Soluciones tecnológicas a problemáticas sociales y ciudadanas. el caso de los retos cívicos y públicos en México. *Teoría de la Educación. Educación y Cultura en la Sociedad de la Información*, 18(1):101–117, 2017. (Citado en página 19.)
- [5] Cira Alvear. Compendio de gestión pública: Normas relativas a la información, probidad y transparencia. https://www.leychile.cl/Consulta/aportar_leychile?itemsporpagina=10&totalitems=20&npagina=1&grupo_aporte=1752&agr=&sub=&comp=3&tipCat=, 2015. Recuperado el 20 de junio del 2017. (Citado en página 23.)
- [6] Anticorruccion.cl. Legislación. http://www.anticorruccion.cl/web_site/appmanager/portal/main?_nfpb=true&_pageLabel=page_Legis1, 2013. Recuperado el 20 de junio del 2017. (Citado en página 23.)
- [7] Joseph P Bailey and Yannis Bakos. An exploratory study of the emerging role of electronic intermediaries. *International Journal of Electronic Commerce*, 1(3):7–20, 1997. (Citado en página 8.)
- [8] C. Batini and M. Scannapieco. *Data Quality: Concepts, Methodologies and Techniques*. Data-Centric Systems and Applications. Springer Berlin Heidelberg, 2006. (Citado en páginas 53 y 55.)
- [9] Carlo Batini, Cinzia Cappiello, Chiara Francalanci, and Andrea Maurino. Methodologies for data quality assessment and improvement. *ACM Comput. Surv.*, 41(3):16:1–16:52, July 2009. (Citado en página 50.)

- [10] T. Berners-Lee. Is your linked open data 5 star? <https://www.w3.org/DesignIssues/LinkedData.html>, 2006. Recuperado el 11 de abril del 2017. (Citado en páginas 15, 47, 50, 52, 53 y 55.)
- [11] Juan Bertón, Patricia Totorica, and Silvia González. Estudio diagnóstico - participación ciudadana electrónica. https://www.agesic.gub.uy/innovaportal/file/3358/1/obsciu_infreferentes_v12.pdf, 2013. Recuperado el 20 de junio del 2017. (Citado en página 25.)
- [12] Christian Bizer, Tom Heath, and Tim Berners-Lee. Linked data-the story so far. *International journal on semantic web and information systems*, 5(3):1–22, 2009. (Citado en página 14.)
- [13] Ismael Caballero, Manuel Serrano, and Mario Piattini. A data quality in use model for big data. In *International Conference on Conceptual Modeling*, pages 65–74. Springer, 2014. (Citado en página 39.)
- [14] Victoria Anderica Caffarena. El valor del activismo de datos en el trabajo de la sociedad civil. *Revista Dígitos*, 1(3):199–220, 2017. (Citado en página 19.)
- [15] Angélica Caro, Coral Calero, Ismael Caballero, and Mario Piattini. A proposal for a set of attributes relevant for web portal data quality. *Software Quality Journal*, 16(4):513–542, Dec 2008. (Citado en páginas 43 y 44.)
- [16] C. Carrasco. Gobierno de datos abiertos en México: Transparencia, rendición de cuentas y escrutinio ciudadano en la era digital. <http://132.248.9.195/ptd2017/enero/305185189/Index.html>, 2016. Recuperado el 11 de abril del 2017. (Citado en página 30.)
- [17] Alejandra Cázares and Guillermo Cejudo. El acceso a la información gubernamental: análisis empírico de políticas de transparencia en cuatro países centroamericanos. *Revista de Gestión Pública*, 2(2):335–381, 2013. (Citado en páginas 30 y 68.)
- [18] Jennifer Arroyo Chacón. Primeros frutos en la lucha contra la corrupción en Costa Rica: condenatoria en sede penal por delitos tipificados por la ley contra la corrupción. *Derecho Penal y Criminología*, 36(101), 2015. (Citado en página 24.)
- [19] Mavis Chan, Peter Johnson, and Malcolm Shookner. Assessing the use of government open data and the role of data intermediaries the case of Nova Scotia’s community counts program. *JeDEM-eJournal of eDemocracy and Open Government*, 8(1):1–27, 2016. (Citado en página 15.)
- [20] Min Chen, David Ebert, Hans Hagen, Robert S Laramee, Robert Van Liere, Kwan-Liu Ma, William Ribarsky, Gerik Scheuermann, and Deborah Silver. Data, information, and knowledge in visualization. *IEEE Computer Graphics and Applications*, 29(1), 2009. (Citado en páginas 6 y 36.)
- [21] S. Chesterman. *Data Protection Law in Singapore: Privacy and Sovereignty in an Interconnected World*. Monograph Series. Academy Publishing, 2014. (Citado en

página 8.)

- [22] A. Chiaretti. Datos abiertos (enlazados) y democratización del acceso a la información en Chile. http://eprints.rclis.org/18677/1/Chiaretti_Simposio14_2013.pdf, 2013. Recuperado el 11 de abril del 2017. (Citado en páginas 14 y 17.)
- [23] Civic Analytics Network. 8 guidelines to help deliver on the promise of a transparent government. <http://www.govtech.com/data/8-Guidelines-to-Help-Deliver-on-the-Promise-of-a-Transparent-Government.html>, 2017. Recuperado el 11 de abril del 2017. (Citado en página 17.)
- [24] H. Collins. *Corporate Portal Definition and Features*. AMACOM, 2001. (Citado en página 43.)
- [25] Joint Research Centre-European Commission et al. *Handbook on constructing composite indicators: Methodology and user guide*. OECD publishing, 2008. (Citado en páginas 55 y 56.)
- [26] Gastón Concha and Alejandra Naser. El desafío hacia el gobierno abierto en la hora de la igualdad. *Comisión Económica para América Latina y el Caribe (CEPAL)*, 2012. (Citado en página 18.)
- [27] J. Córdoba. Compendio de leyes costarricenses sobre acceso a la información pública y transparencia. http://www.dhr.go.cr/Red_de_transparencia/indice_transparencia/jurisprudencia/compendio_acceso_informacion_2013.pdf, 2013. Recuperado el 11 de abril del 2017. (Citado en página 30.)
- [28] A. Cruz, S. Ettinger, I. Pousadela, and G. Correa. El poder de la transparencia fontra: un programa de la sociedad civil del cono sur. http://www.rendircuentas.org/wp-content/uploads/2012/03/Libro_Fontra.pdf, 2012. Recuperado el 11 de abril del 2017. (Citado en páginas 18, 21, 22 y 30.)
- [29] César Nicandro Cruz-Rubio. ¿qué es (y que no es) gobierno abierto? una discusión conceptual. *EUNOMÍA. Revista en Cultura de la Legalidad*, pages 37–53, 2015. (Citado en página 17.)
- [30] Piet JH Daas, Judit Arends-Tóth, Barry Schouten, and Léander Kuijvenhoven. Proposal for a quality framework for the evaluation of administrative and survey data. In *Paper for the workshop on the Combination of surveys and administrative data*, pages 29–30, 2008. (Citado en páginas 36 y 47.)
- [31] Tim Davies. Open data, democracy and public sector reform. <http://www.opendataimpacts.net/report/wp-content/uploads/2010/08/How-is-open-government-data-being-used-in-practice.pdf>, 2010. Recuperado el 10 de mayo del 2017. (Citado en página 15.)
- [32] Unidad de Modernización y Gobierno Digital. Norma técnica para publicación de datos abiertos en Chile. Reporte técnico, Gobierno de Chile, 2013. Recuperado el 11 de abril del 2017 de: <http://instituciones.gobiernoabierto.cl/>

[NormaTecnicaPublicacionDatosChile_v2-1.pdf](#). (Citado en páginas 54, 55, 63, 65 y 74.)

- [33] Consejo de Transparencia. Índice de transparencia y acceso a la información: Construcción del modelo y primeros resultados. http://www.consejotransparencia.cl/consejo/site/artic/20121213/asocfile/20121213160518/indice_de_transparencia_y_acceso_a_la_informacion.pdf, 2013. Recuperado el 11 de abril del 2017. (Citado en páginas 62, 63 y 64.)
- [34] David Dufty, Hélène Bérard, Laurie Reedman, Sylvie Lefranc, Marina Signore, Juan Muñoz, Enrique Ordaz, Jacek Maślankowski, Dominik Rozkrut, Boro Nikic, Ronald Jansen, and Karoly Kovacs. A suggested framework for the quality of big data. Reporte técnico, Oficina nacional de estadística, Canadá, 2015. <http://www.fields.utoronto.ca/talks/suggested-framework-national-statistics-offices-assessing-quality-big-data>. (Citado en páginas 39, 40, 52 y 55.)
- [35] Suruchi Dumpawar. *Open government data intermediaries: Mediating data to drive changes in the built environment*. PhD thesis, Massachusetts Institute of Technology, 2015. (Citado en página 16.)
- [36] David Eaves. The three laws of open government data. <https://eaves.ca/2009/11/29/three-laws-of-open-data-international-edition/>, 2009. Recuperado el 20 de junio del 2017. (Citado en página 45.)
- [37] Larry P. English. *Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits*. John Wiley & Sons, Inc., New York, NY, USA, 1999. (Citado en páginas 50, 53 y 55.)
- [38] A. Fierro, A. García, A. Ríos, R. Velázquez, and D. Zavala. Medición de transparencia y acceso a la información en México. http://www.metricadetransparencia.cide.edu/download.php?file=Reporte_Nacional_20150827.pdf, 2014. Recuperado el 11 de abril del 2017. (Citado en páginas 29, 30, 47, 54, 55 y 68.)
- [39] Organisation for Economic Co-operation and Development. *OECD Principles and Guidelines for Access to Research Data from Public Funding*. OECD Pub., 2007. (Citado en páginas 12, 13, 54, 55, 56 y 57.)
- [40] Christopher Fox, Anany Levitin, and Thomas Redman. The notion of data and its quality dimensions. *Information processing & management*, 30(1):9–19, 1994. (Citado en páginas 39 y 50.)
- [41] Freedomdefined. Definition. <http://freedomdefined.org/Definition/Es>, 2009. Recuperado el 11 de abril del 2017. (Citado en página 9.)
- [42] Guía digital. Decretos supremos. <http://www.guiadigital.gob.cl/articulo/decretos-supremos>, 2015. Recuperado el 20 de junio del 2017. (Citado en página 23.)

- [43] Luis A Guerrero and José A Pino. Understanding organizational memory. In *Computer Science Society, 2001. SCCC'01. Proceedings. XXI International Conference of the Chilean*, pages 124–132. IEEE, 2001. (Citado en página 7.)
- [44] John Hagel, Jeffrey F Rayport, et al. The new infomediaries. *The McKinsey Quarterly*, 1(4):54, 1997. (Citado en página 8.)
- [45] Teresa M Harrison, Theresa A Pardo, and Meghan Cook. Creating open government ecosystems: A research and development agenda. *Future Internet*, 4(4):900–928, 2012. (Citado en página 8.)
- [46] M. Hausenblas. Las cinco estrellas de los datos abiertos. <http://5stardata.info/es/>, 2012. Recuperado el 11 de abril del 2017. (Citado en página 16.)
- [47] Jinwon Ho and Rong Tang. Towards an optimal resolution to information overload: An infomediary approach. In *Proceedings of the 2001 International ACM SIGGROUP Conference on Supporting Group Work, GROUP '01*, pages 91–96, New York, NY, USA, 2001. ACM. (Citado en página 8.)
- [48] Marijn Janssen and Anneke Zuiderwijk. Infomediary business models for connecting open data providers and users. *Social Science Computer Review*, 32(5):694–711, 2014. (Citado en página 8.)
- [49] Marijn Janssen and Anneke Zuiderwijk. Infomediary business models for connecting open data providers and users. *Social Science Computer Review*, 32(5):694–711, 2014. (Citado en página 15.)
- [50] Alan F. Karr, Ashish P. Sanil, and David L. Banks. Data quality: A statistical perspective. *Statistical Methodology*, 3(2):137 – 173, 2006. (Citado en páginas 36 y 47.)
- [51] E. Lanza and M. Mas. Índice de transparencia activa en línea, el estado uruguayo y la provisión de información pública a través de la web. <http://itael.cainfo.org.uy/wp-content/uploads/2016/08/ITaEL2015-.pdf>, 2015. Recuperado el 11 de abril del 2017. (Citado en páginas 47, 54 y 55.)
- [52] Nuno Laranjeiro, Seyma Nur Soydemir, and Jorge Bernardino. A survey on data quality: classifying poor data. In *Dependable Computing (PRDC), 2015 IEEE 21st Pacific Rim International Symposium on*, pages 179–188. IEEE, 2015. (Citado en páginas 41, 42, 48 y 171.)
- [53] Anany Levitin and Thomas Redman. Quality dimensions of a conceptual view. *Information Processing & Management*, 31(1):81–88, 1995. (Citado en página 39.)
- [54] Juho Lindman, Matti Rossi, and Virpi Kristiina Tuunainen. Open data services: Research agenda. In *System Sciences (HICSS), 2013 46th Hawaii International Conference on*, pages 1239–1246. IEEE, 2013. (Citado en página 15.)
- [55] Jennifer Long and Craig Seko. A new method for database data quality evaluation at the canadian institute for health information (cihi). In *IQ*, pages 238–250, 2002.

(Citado en páginas 39, 50 y 52.)

- [56] David Loshin. 2 - who owns information? In David Loshin, editor, *Enterprise Knowledge Management*, The Morgan Kaufmann Series in Data Management Systems, pages 25 – 46. Academic Press, San Diego, 2001. (Citado en páginas 50, 52, 53, 55 y 56.)
- [57] Fritz Machlup. *Knowledge: its creation, distribution and economic significance, Volume I: Knowledge and knowledge production*, volume 1. Princeton University Press, 2014. (Citado en página 6.)
- [58] Fritz Machlup. *Knowledge: Its Creation, Distribution and Economic Significance, Volume II: The Branches of Learning*, volume 2. Princeton university press, 2014. (Citado en página 6.)
- [59] Fritz Machlup. *Knowledge: Its Creation, Distribution and Economic Significance, Volume III: The Economics of Information and Human Capital*, volume 3. Princeton University Press, 2014. (Citado en página 6.)
- [60] Gustavo Magalhaes, Catarina Roseira, and Sharon Strover. Open government data intermediaries: A terminology framework. In *Proceedings of the 7th International Conference on Theory and Practice of Electronic Governance*, pages 330–333. ACM, 2013. (Citado en página 16.)
- [61] Francis Maude. Open data white paper-unleashing the potential. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/78946/CM8353_acc.pdf, 2012. Recuperado el 12 de mayo del 2017. (Citado en páginas 11, 50, 53, 54 y 55.)
- [62] Eunice Mercado-Lara and José Ramón Gil García. Gobierno abierto e intermediarios de datos: El caso de aiddata. https://www.researchgate.net/profile/Eunice_Mercado-Lara/publication/281444488_Gobierno_abierto_e_intermediarios_de_datos_El_caso_de_AidData/links/55e7830b08ae3e1218420e1c.pdf, 2014. Recuperado el 11 de abril del 2017. (Citado en página 16.)
- [63] Luisa Montuschi. Datos, información y conocimiento. de la sociedad de la información a la sociedad del conocimiento. *Serie Documentos de Trabajo de la Universidad del CEMA*, 192, 2001. (Citado en página 6.)
- [64] E. Mora, P. Guzmán, and R. Herrera. Documento metodológico. y su sitio web. . . ¿es transparente? guía de diseño para portales web transparentes. <http://cicap.ucr.ac.cr/web/guia-de-transparencia/>, 2017. Recuperado el 11 de abril del 2017. (Citado en páginas 29 y 47.)
- [65] C. Moraga, M. Á. Moraga, C. Calero, and A. Caro. Square-aligned data quality model for web portals. In *2009 Ninth International Conference on Quality Software*, pages 117–122, Aug 2009. (Citado en páginas 43, 45 y 46.)
- [66] Carmen Moraga, Ma Ángeles Moraga, Coral Calero, and Ángélica Caro. Towards the discovery of data quality attributes for web portals. In Martin Gaedke, Michael Gross-

- niklaus, and Oscar Díaz, editors, *Web Engineering*, pages 251–259, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg. (Citado en página 44.)
- [67] B. Nardi and V. O’Day. *Information Ecologies: Using Technology with Heart*, 1999. <http://firstmonday.org/ojs/index.php/fm/article/view/672/582>. (Citado en página 8.)
- [68] National Science Foundation. Open Data at NSF. <https://www.nsf.gov/data/>, 2013. Recuperado el 20 de junio del 2017. (Citado en páginas 13 y 50.)
- [69] Felix Naumann and Claudia Rolker. Assessment methods for information quality criteria. In *In Proceedings of the International Conference on Information Quality (IQ, 2000*. (Citado en páginas 39, 52, 53, 54, 55, 56 y 57.)
- [70] ONU. Encuestas e-gobierno de la onu. <http://workspace.unpan.org/sites/Internet/Documents/UNPAN96407.pdf>, 2016. Recuperado el 20 de junio del 2017. (Citado en página 20.)
- [71] Open Knowledge International. The open data handbook. <http://opendatahandbook.org/guide/es>, 2007. Recuperado el 11 de abril del 2017. (Citado en página 10.)
- [72] Open Knowledge International. Open definition 2.1. <http://opendefinition.org/od/2.0/es/>, 2017. Recuperado el 11 de abril del 2017. (Citado en páginas 9 y 10.)
- [73] opendatacharter. Carta internacional de datos abiertos. <http://opendatacharter.net/principles-es/>, 2015. Recuperado el 11 de abril del 2017. (Citado en páginas 11, 13, 54 y 55.)
- [74] Opengovdata. Principios para los datos de gobierno abierto. https://public.resource.org/8_principles.html, 2007. Recuperado el 12 de mayo del 2017. (Citado en páginas 10, 11, 13, 45, 50, 53 y 55.)
- [75] Tim O’Reilly. Government as a platform. *innovations*, 6(1):13–40, 2010. http://www.mitpressjournals.org/doi/pdf/10.1162/INOV_a_00056. (Citado en página 8.)
- [76] Seán O’Riain, Edward Curry, and Andreas Harth. Xbrl and open data for global financial ecosystems: A linked data approach. *International Journal of Accounting Information Systems*, 13(2):141–162, 2012. (Citado en página 8.)
- [77] Edgar Oviedo Blanco. *Modelo de madurez para portales de datos abiertos e incorporación a la norma técnica nacional de Costa Rica*. PhD thesis, Universitat d’Alacant-Universidad de Alicante, 2016. <http://hdl.handle.net/10045/54316>. (Citado en página 47.)
- [78] Jorge Alejandro Patiño. Datos abiertos y ciudades inteligentes en américa latina. *Estudio de casos (Open data and smart cities in Latin America. Case studies) Comisión Económica para América Latina y el Caribe (CEPAL)*, Santiago de Chile, 2014. (Citado en página 25.)

- [79] Alejandro Prince, Lucas Jolias, and Carlos Brys. Análisis de la cadena de valor del ecosistema de datos abiertos de la ciudad de buenos aires. In *Ponencia presentada en Conferencia Regional de Datos Abiertos para América Latina y el Caribe, Montevideo*, 2013. (Citado en página 8.)
- [80] T.C. Redman. *Data Quality: Management and Technology*. Bantam professional books. Bantam Books, 1992. (Citado en página 39.)
- [81] T.C. Redman. *Data Quality: The Field Guide*. Books 24x7 IT PRO. Digital Press, 2001. (Citado en páginas 39, 43 y 52.)
- [82] Thomas C. Redman. *Data Quality for the Information Age*. Artech House, Inc., Norwood, MA, USA, 1st edition, 1997. (Citado en páginas 39 y 52.)
- [83] Luigi Reggi and Sharon Dawes. Open government data ecosystems: Linking transparency for innovation with transparency for participation and accountability. In *International Conference on Electronic Government and the Information Systems Perspective*, pages 74–86. Springer, 2016. (Citado en página 8.)
- [84] Shazia Sadiq. *Handbook of data quality: research and practice*. Springer, 2013. (Citado en páginas 39, 40 y 50.)
- [85] Igbal Safarov, Albert Meijer, and Stephan Grimmelikhuijsen. Utilization of open government data: A systematic literature review of types, conditions, effects and users. *Information Polity*, 22(1):1–24, 2017. <https://doi.org/10.3233/IP-160012>. (Citado en página 31.)
- [86] Rodrigo Sandoval-Almazán. Gobierno abierto y transparencia: construyendo un marco conceptual. *Convergencia*, 22(68):203–227, 2015. (Citado en páginas 16 y 17.)
- [87] Francois Van Schalkwyk, Michael Canares, Sumandro Chattapadhyay, and Alexander Andrason. Open Data Intermediaries in Developing Countries. https://figshare.com/articles/Open_Data_Intermediaries_in_Developing_Countries/1449222, 2015. Recuperado el 11 de abril del 2017. (Citado en página 16.)
- [88] Pablo Schiavi. El desafío de los gobiernos municipales en el uruguay: participación ciudadana, transparencia y articulación. *Revista de Derecho*, 1(22):81–106, 2012. <http://hdl.handle.net/2183/12051>. (Citado en página 30.)
- [89] Andrew Schrock and Gwen Shaffer. Data ideologies of an interested public: A study of grassroots open government data intermediaries. *Big Data & Society*, 4(1):2053951717690750, 2017. (Citado en página 16.)
- [90] Javier Pérez Seveso, Mariana Mas, Matías Ponce, Giorgina Santangelo, and Mauricio Coitiño. Índice de transparencia activa en línea 2015. <http://itael.cainfo.org.uy/wp-content/uploads/2016/08/ITaEL2015-.pdf>, 2015. Recuperado el 11 de abril del 2017. (Citado en página 29.)
- [91] Sebastián Soto Velasco. Ley de transparencia desafíos en su aplicación. *Temas de*

- la Agenda Pública*, 2009. <https://repositorio.uc.cl/bitstream/handle/11534/1528/531811.pdf?sequence=1&isAllowed=y>. (Citado en página 30.)
- [92] J. Talledo. *Acceso a datos en aplicaciones web del entorno servidor*. España: Paraninfo, 2016. (Citado en página 6.)
- [93] Francis N Teskey. User models and world models for data, information, and knowledge. *Information processing & management*, 25(1):7–14, 1989. (Citado en página 6.)
- [94] Sarah Treuhaft. The democratization of data: How the internet is shaping the work og data intermediaries. Reporte técnico, Working paper//Institute of Urban and Regional Development, 2006. <https://escholarship.org/content/qt32961226/qt32961226.pdf?nosplash=f320c6dab04706e724a2fd58d4bd48f9>. (Citado en página 8.)
- [95] U.S Government. Transparency and Open Government. <https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government>, 2009. Recuperado el 20 de junio del 2017. (Citado en página 16.)
- [96] L. Velázquez. *Desarrollo de Habilidades en el uso de las Tecnologías de la Información y la Comunicación*. Mexico: Velazquez, 2016. (Citado en página 6.)
- [97] Antonio Vetrò, Lorenzo Canova, Marco Torchiano, Camilo Orozco Minotas, Raimondo Iemma, and Federico Morando. Open data quality measurement framework: Definition and application to open government data. *Government Information Quarterly*, 33(2):325 – 337, 2016. (Citado en páginas 46, 48 y 49.)
- [98] Pablo Viollier. El estado de la protección de datos personales en chile. <https://www.derechosdigitales.org/wp-content/uploads/PVB-datos-int.pdf>, 2017. Recuperado el 20 de junio del 2017. (Citado en página 19.)
- [99] Yair Wand and Richard Y Wang. Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, 39(11):86–95, 1996. (Citado en páginas 39 y 50.)
- [100] Richard Y Wang and Diane M Strong. Beyond accuracy: What data quality means to data consumers. *Journal of management information systems*, 12(4):5–33, 1996. (Citado en páginas 37, 38, 39, 52, 55, 56 y 57.)
- [101] Y. Richard Wang and Lisa M. Guarascio. Dimensions of data quality: Toward quality data by design. <http://web.mit.edu/smadnick/www/wp2/1991-06.pdf>, 1991. Recuperado el 11 de abril del 2017. (Citado en páginas 38 y 39.)
- [102] D. Zamora. Documento metodológico del Índice de transparencia del sector público costarricense, basado en sitios web. http://www.dhr.go.cr/red_de_transparencia/indice_de_transparencia_del_sector_publico.aspx, 2015. Recuperado el 11 de abril del 2017. (Citado en páginas 29, 63 y 66.)

Anexos

Anexos A

Cuestionario aplicado a los sitios

Plantilla

Este formulario corresponde al portal de datos abiertos de (nombreMinistero). Las páginas fuente fueron:

<http://www.paginafuente.com>

*Obligatorio

1. Hipervínculo *

2. Nombre del archivo *

1-Compleitud

3. ¿Se encuentra el conjunto de datos completo, es decir, no tiene ningún dato ausente en sus registros?

Marca solo un óvalo.

Sí

No

2-Procesable por máquina

4. ¿El archivo donde se encuentran los datos tiene una extensión o formato que facilita la automatización de la manipulación de los datos?

Marca solo un óvalo.

Sí

No

5. ¿Posee el conjunto de datos en su contenido interno registros estructurados?

Marca solo un óvalo.

Sí

No

3-Primario

6. ¿Los datos se encuentran en un grado fino de granularidad, sin cálculos ni agregaciones?

Marca solo un óvalo.

Sí

No

4-Licenciamiento

7. ¿El sitio indica la licencia de los datos?

Marca solo un óvalo.

- No indica licencia
- Open Data Commons Public Domain Dedication and License
- Reconocimiento
- Reconocimiento – Sin Derivar
- Reconocimiento – Sin Derivar – No comercial
- Reconocimiento – No comercial
- Reconocimiento – No comercial – Compartir igual
- Reconocimiento – Compartir igual
- copyright
- Otros: _____

8. ¿El formato del archivo donde se almacenan los datos posee una licencia que permite acceder a ellos sin costo/abierto?

Marca solo un óvalo.

- Sí
- No

5-Consistencia

De las siguientes pre

9. ¿El nombre de los atributos es igual en las diferentes entregas que se realizan periódicamente?

Marca solo un óvalo.

- Sí
- No

10. ¿El tipo de datos es siempre el mismo? Por ejemplo, siempre es entero

Marca solo un óvalo.

- Sí
- No

11. ¿El tamaño de los datos es siempre el mismo?

Marca solo un óvalo.

- Sí
- No

12. ¿La estructura de los datos es siempre la misma?

Marca solo un óvalo.

- Sí
- No

Consistencia final

Si alguna de las preguntas anteriores obtuvo un NO como respuesta, elija No en caso contrario Sí

13. **¿Es consistente? ***

Marca solo un óvalo.

Sí

No

6-Coherencia

14. **¿En el conjunto de datos los valores de los datos de un mismo campo están ajustados a una medida única determinada?**

Marca solo un óvalo.

Sí

No

7-Metadatos

15. **¿El conjunto de datos posee los metadatos definidos como "metadatos mínimos"?**

Marca solo un óvalo.

Sí

No

8-Documentación

16. **¿Posee el conjunto de datos al menos un archivo con definiciones y/o explicaciones que ayuden al usuario a interpretar los datos?**

Marca solo un óvalo.

Sí

No

9-Descripción

17. **¿Posee el sistema donde se publica el conjunto de datos una descripción del contenido del conjunto?**

Marca solo un óvalo.

Sí

No

10-Privacidad

18. **¿Los datos publicados respetan la legislación vigente del país respecto al tema de privacidad?**

Marca solo un óvalo.

Sí

No

11-Estandarizado

19. **¿En el conjunto de datos los valores de los datos de un mismo campo están ajustados a una medida única la cual es exacta y pertenece al sistema internacional de unidades?**

Marca solo un óvalo.

Sí

No

12-Sostenibilidad

20. **¿El conjunto de datos se publica de forma periódica?**

Marca solo un óvalo.

Sí

No

21. **¿Existe un compromiso de mantener la publicación del conjunto de datos en forma periódica?**

Marca solo un óvalo.

Sí

No

13-Seguridad

22. **¿Posee el sistema donde se alojan los datos mecanismos para agregar, modificar o eliminar conjuntos de datos con el nivel de seguridad correcto, es decir, que realice las validaciones de usuario pertinentes de manera que no ponga en riesgo la calidad de los datos?**

Marca solo un óvalo.

Sí

No

14-Portabilidad

23. **¿El sistema donde se alojan los datos ofrece la posibilidad de compartir, descargar los datos en bulto o través de una API?**

Marca solo un óvalo.

Sí

No

15-Capacidad de relacionarse

24. **¿El conjunto de datos se encuentra asociado a un identificador de recursos uniforme, ya sea a través de metadatos o en una etiqueta en el sistema?**

Marca solo un óvalo.

Sí

No

16-Relevancia/importancia

25. ¿Los datos son importantes/relevantes para el usuario?

Marca solo un óvalo.

- Sí
 No

17-Pertinencia

26. ¿Los datos son pertinentes para la necesidad del usuario?

Marca solo un óvalo.

- Sí
 No

18-Actualizado

27. Los datos publicados demuestran que la actualización del conjunto de datos se entrega según:

Marca solo un óvalo.

- La forma que indica la ley del país
 La forma que indica el portal
 Los datos cambian en el mundo real
 Otra forma
 No se actualizan
 Anualmente
 mensualmente

19-Fácil de entender/Interpretable/Comprensible

28. ¿Los datos se pueden interpretar con facilidad?

Marca solo un óvalo.

- Sí
 No

20-Claridad contextual

29. ¿El sitio provee los datos suficientes para tener claro el contexto de los datos?

Marca solo un óvalo.

- Sí
 No

21-Contactable

30. ¿El sistema ofrece la posibilidad de comunicarse con el responsable de los datos?

Marca solo un óvalo.

- Sí
 No

31. **¿El sistema brinda los datos suficientes para establecer comunicación con el responsable de los de datos?**

Marca solo un óvalo.

- Sí
 No

22-Catálogo de datos

32. **¿Posee el sitio un catálogo de datos?**

Marca solo un óvalo.

- Sí
 No

23-Verificable/trazable

33. **¿Posee el sitio los datos sobre del conjunto de datos de manera que se puedan realizar operaciones de auditoría?**

Marca solo un óvalo.

- Sí
 No

24-Volumen de datos

34. **¿En qué nivel se encuentra el volumen de datos del conjunto de datos?**

Marca solo un óvalo.

- Fácil de manejar
 Manejable
 Difícil de manejar
 Intratable

25-Almacenamiento/custodia

35. **¿Posee el conjunto de datos o el sistema el nombre de la organización responsable de almacenar o custodiar los datos de forma permanente?**

Marca solo un óvalo.

- Sí
 No

26-Observaciones

36. **Observaciones**

Anexos B

Imágenes de portales de datos abiertos

MEC | Publicaciones x

educacion.mec.gub.uy/innovaportal/v/11078/5/mecweb/publicaciones_?3colid=927&breadid=927

mec Dirección de Educación
MINISTERIO DE EDUCACIÓN Y CULTURA URUGUAY

Transparencia Convocatorias Adquisiciones Preguntas frecuentes Contacto

Buscar

Inicio · Educación · Cultura · Centros MEC · D2C2 · Cooperación Internacional · Asuntos Constitucionales, Legales y Registrales

Acerca de

Programas y proyectos transversales

Primera infancia

Educación superior

Educación no formal

Trámites

Gestión

Comisiones y redes
Investigación y estadística
Documentación pedagógica
Asuntos internacionales

Comunidad

Ajedrez para la convivencia
Cultura Científica
Plantar es cultura
Talleres en el liceo

Registros y habilitaciones

Habilitación de Escuelas Privadas de Enfermería
Habilitación al Sistema Nacional Integrado de Cuidados
Instituciones educativas autorizadas y habilitadas

Investigación y estadística

- Investigación
- Estadística
- Reportes internacionales
- Sitios de interés
- Solicitud de información

Publicaciones

Panorama de la Educación 2016



[Panorama-2016.xlsm](#)

[Presentación Panorama 2016 \(pptx\)](#)

ANUARIO ESTADÍSTICO DE EDUCACIÓN 2016



[Anuario Estadístico de Educación 2016 \(.xls\)](#)

[Indicadores ODS 4 Uruguay.xlsx](#)

[ODS 4. Aproximación de la Metodología de Naciones Unidas - Uruguay.pdf](#)

[Presentación ODS.pptx](#)

f t G+

Me gusta 0

GobiernoTransparente - x

portales.mineduc.cl/transparencia/inicio.html

Última actualización: 12/04/2018

<p>Marco Normativo</p> <p>Marco normativo aplicable</p> <p>Potestades, competencias, responsabilidades, funciones, atribuciones y/o tareas</p> <p>Actos y documentos publicados en el Diario Oficial</p>	<p>Compras y Adquisiciones</p> <p>En Mercado Público</p> <p>Otras compras y adquisiciones</p>	<p>Trámites del Organismo</p> <p>Trámites en ChileAtiende</p> <p>Otros trámites</p>	<p>Derecho de Acceso a la Información Pública</p> <p>Solicitud de Información Ley de Transparencia</p> <p>Índice de actos y documentos calificados como secretos o reservados</p>
<p>Actos y Resoluciones</p> <p>Actos con efectos sobre terceros</p>	<p>Transferencias</p> <p>Registro de la ley N° 19.862</p> <p>Otras transferencias</p>	<p>Participación Ciudadana</p> <p>Acceso al Portal de Inclusión y Participación Ciudadana</p> <p>Mecanismos de participación ciudadana</p> <p>Norma General de Participación Ciudadana</p> <p>Resolución Exenta N°4718 de 2015 modifica Norma de Participación Ciudadana</p>	<p>Costos Directos de Reproducción</p> <p>Costos directos de reproducción de la información requerida via ley N° 20.285</p>
<p>Estructura Orgánica</p> <p>Facultades, funciones y atribuciones de sus unidades u órganos internos</p> <p>Diagrama de la estructura orgánica</p>	<p>Información Presupuestaria</p> <p>Informes de ejecución Presupuestaria</p>	<p>Subsidios y Beneficios</p> <p>Programas de subsidios y otros beneficios</p> <p>Nómina de beneficiarios de programas sociales</p>	<p>Declaración de patrimonio e intereses</p> <p>Declaración de patrimonio de intereses Ley N° 20.880</p> <p>Información histórica sobre declaración de patrimonio e intereses</p>
<p>Dotación de Personal</p> <p>Dotación de planta</p> <p>Dotación a contrata</p> <p>Remuneraciones según estamentos, grados o jornadas</p> <p>Dotación a honorarios</p> <p>Otros contratos sujetos al Código del Trabajo</p>	<p>Auditorías al Ejercicio Presupuestario</p> <p>Auditorías al ejercicio presupuestario</p>	<p>Vínculos Institucionales</p> <p>Participación, representación e intervención</p>	<p>Información ley N° 20.285</p> <p>Normas de la Ley sobre Acceso a la Información Pública</p> <p>Tutoriales Derecho de Acceso</p>
	<p>Datos de Interés Ciudadano</p> <p>Datos de Interés Ciudadano</p>	<p>Antecedentes preparatorios de normas que afecten a EMT Ley N°20.416</p> <p>Formularios y antecedentes de propuestas normativas que afecten a empresas de menor tamaño (EMT)</p>	<p>Ley del Lobby</p> <p>Acceso a Subsitio Lobby</p> <p>Resolución Exenta N° 1992 del 27 de Abril de 2017 que designa a los funcionarios a considerar como sujetos pasivos en conformidad a la Ley N° 20.730</p>

imprimir

Anexos C

Literatura de dimensiones de datos

Estas dos figuras son parte del trabajo realizado por Laranjeiro y colegas [\[52\]](#)

Research work	Structure	Terms used
Beyond accuracy: What data quality means to data consumers [15]	4 categories, 15 dimensions	Intrinsic: accuracy, objectivity, believability, reputation Contextual: value-added, relevancy, timeliness, completeness, appropriate amount of data Representational: interpretability, ease of understanding, representational consistency, concise representation Accessible: accessibility, access security
Anchoring data quality dimensions in ontological foundations [31]	2 views, 2 categories, 26 dimensions	Internal view Data-related: accuracy, reliability, timeliness, completeness, currency, consistency, precision System-related: reliability External view Data-related: timeliness, relevance, content, importance, sufficiency, usability, usefulness, clarity, conciseness, freedom from bias, informativeness, level of detail, quantitiveness, scope, interpretability, understandability System-related: timeliness, flexibility, format, efficiency
DoD Guidelines... [26]	6 characteristics	Accuracy, completeness, consistency, timeliness, uniqueness, validity
Data Quality Handbook... [42]	5 dimensions	Correctness, completeness, consistency, currency, accessibility
Data Quality for the Information Age [32]	3 categories, 27 dimensions (15 characteristics)	Data model: content (data relevance, obtainability of values, clarity of definition), scope (comprehensiveness, essentialness), level of detail (attribute granularity, domain precision), composition (naturalness, identifiability, homogeneity, minimal unnecessary redundancy), model consistency (structural, semantic consistency), reaction to change (flexibility, robustness) Data values: accuracy, completeness, currency, and consistency Data representation: appropriateness, interpretability, portability, format precision, format flexibility, ability to represent null values, efficient usage of recording media, representation consistency
Improving data warehouse and business information quality [33]	2 categories, 15 dimensions	Inherent: definition conformance, value completeness, validity/business rule conformance, accuracy to surrogate source, accuracy to reality, precision, non-duplication, equivalence/concurrency of redundant or distributed data Pragmatic: accessibility, timeliness, contextual clarity, derivation integrity, usability, rightness or fact completeness
Assessment methods for information quality criteria [34]	3 classes, 22 criteria	Subject: believability, concise representation, interpretability, relevancy, reputation, understandability, value-added Object: completeness, customer support, documentation, objectivity, price, reliability, security, timeliness, verifiability Process: accuracy, amount of data, availability, consistent representation, latency, response time Accessibility/Delivery: availability, protocol, security
Data Quality: The Guide Field [41]	9 categories, 51 dimensions	Quality of content: attribute granularity, comprehensiveness, essentialness, flexibility, appropriate use, areas covered, homogeneity, naturalness, obtainability, precision of domains, robustness, semantic consistency, structural consistency, simplicity, clear definition, identifiability, source, relevancy Quality of values: accuracy, completeness, timeliness, consistency Presentation quality: appropriateness, format precision, use of storage Flexibility: portability, representation consistency, null values, formats, language, ease of interpretation Improvement: feedback, measurement, track record Privacy: consumer privacy, privacy of others, security Commitment: warning, help, special requests, commitment Architecture: library/documentation, logical structure, physical structure, naming, rules, redundancy, unit cost
Enterprise Knowledge Management: The Data Quality Approach [36]	5 categories, 31 characteristics	Data models: clarity, comprehensiveness, flexibility, robustness, essentialness attribute granularity, domain precision, homogeneity, naturalness, identifiability, obtainability, relevance, simplicity, semantic/structural consistency Data values: accuracy, null values, completeness, consistency, currency/timeliness Information domains: enterprise agreement of usage, stewardship, ubiquity Data presentation: appropriateness, correct interpretation, flexibility, format precision, portability, representation consistency, representation of null values, use of storage Information policy: accessibility, metadata, privacy, redundancy, security, unit cost
Data quality assessment [12]	16 dimensions	Accessibility, appropriate amount of data, believability, completeness, concise/consistent representation, ease of manipulation, free-of-error, interpretability, objectivity, relevancy, reputation, security, timeliness, understandability, value-added
A New Method for Database Data Quality Evaluation at the Canadian Institute... [37]	5 categories 24 characteristics	Accuracy, timeliness, comparability, usability, relevance Over-coverage, under-coverage, simple response variance, reliability, correlated response variance, collection and capture, unit and item non-response, edit and imputation, processing, estimation, timeliness, comprehensiveness, integration, standardization, equivalence, linkage-ability, product comparability, historical comparability, accessibility, documentation, interpretability, adaptability, value

Research work	Structure	Terms used	
AIMQ: a methodology for information quality assessment [38]	4 quadrants (2 views plus 2 views)	Product quality	Conforms to specifications (Sound): free-of-error, concise representation, completeness, consistent representation Meets or exceeds consumer expectations (Useful): Appropriate amount, relevancy, understandability, interpretability, objectivity
		Service Quality	Conforms to specifications (Dependable): Timeliness, security Meets or exceeds consumer expectations (Usable): Believability, accessibility, ease of operation, reputation
A conceptual framework and belief-function approach to assessing overall information quality [39]	4 attributes, 8 subattributes	Accessibility	
		Interpretability: Intelligible, meaningful	
		Relevance: user-specified criteria, timely (considers age and volatility)	
		Credibility: accuracy, completeness, consistency, and non-fictitiousness	
A Framework for Information Quality Assessment [40]	3 categories, 22 dimensions	Intrinsic: Accuracy/validity, cohesiveness, complexity, semantic consistency, structural consistency, currency, informativeness / redundancy, naturalness, precision / completeness	
		Relational or contextual: Accuracy, accessibility, complexity, naturalness, informativeness / redundancy, relevance (aboutness), precision / completeness, security, semantic consistency, structural consistency, verifiability, volatility	
		Reputation: authority	
ISO/IEC 25012 [8]	2 points of view and 1 mixed view, 15 characteristics	Inherent: accuracy, completeness, consistency, credibility, currentness; accessibility, compliance, confidentiality, efficiency, precision, traceability, understandability	
		System dependent: availability, portability, recoverability	
		Inherent and system dependent: accessibility, compliance, confidentiality, efficiency, precision, traceability, understandability	
SPDQM: SQuARE-aligned portal data quality model [43]	2 points of view, 4 categories, 27 characteristics, 15 subcharacteristics	Inherent	Intrinsic: accuracy, credibility (objectivity, reputation), traceability, currentness, expiration, completeness, consistency, accessibility (interactive, operation ease, cust. support), compliance, confidentiality, efficiency, precision, understandability
			Operational: availability, accessibility, verifiability, confidentiality, portability, recoverability
		System dependent	Contextual: validity (reliability, scope), value-added (applicability, flexibility, novelty), relevancy (novelty, timeliness), specialization, usefulness, efficiency, effectiveness, traceability, compliance, precision
			Representational: concise representation, consistent representation, understandability, (interpretability, amount of data, documentation, organization), attractiveness, readability
The Practitioner's Guide to Data Quality Improvement [17]	3 categories, 10 Characteristics	Intrinsic: Accuracy, Lineage, Semantic, Structure.	
		Contextual: completeness, consistency, currency, timeliness, reasonableness and identifiability	
		Qualitative dimensions	
		Process pragmatics: ease of operation, ease of navigation, interactivity, flexibility of representation, suitability of representation, access security, accessibility, latency, response time, availability	
Information Quality Dimensions for the Social Web [45]	5 categories, 42 criteria	Syntactics: consistency, semantic consistency, structural consistency, conformability, naturalness, integrity	
		User pragmatics: believability, verifiability, amount empirical evidence, reliability, reputation, user-conformability, enjoyability	
		Semantics: informativeness, conciseness, accuracy, objectivity, currency, completeness, cohesiveness, degree of context, maintainability, unambiguous	
		Information pragmatics: understandability, interpretability, usability, efficiency, value add, complexity, relevancy, timeliness, volatility	
Prioritization of data quality dimensions and skills requirements in genome annotation work [46]	5 constructs, 19 dimensions	Accuracy: accuracy, unbiased, believability, traceability	
		Accessibility: accessibility, believability, appropriate amount of information	
		Usefulness: interpretability, understandability, ease of manipulation, consistent representation, value added	
		Relevance: relevant, concise representation, up-to-date, reputation, value added	
		Security: security, traceability	
Towards an ontology... [47]	5 dimensions	Completeness, accuracy, correctness, consistency, and timeliness	
A Suggested Framework for the Quality of Big Data [48]	3 hyperdimensions, 11 dimensions, 3	Source: institutional/business environment, privacy and security	
		Metadata: complexity, accessibility clarity, completeness, usability, time factors, coherence(linkability and consistency), validity	
		Data: accuracy (selectivity), coherence (consistency and linkability), validity, time factors	
A Data Quality in Use Model for Big Data [49]	3 by 3 views, 15 characteristics	Contextual consistency	Velocity: consistency, credibility, confidentiality Volume: completeness, credibility Variety: accuracy, consistency, understandability
		Temporal	Velocity: consistency, credibility, currentness, availability Volume: availability Variety: consistency, currentness, compliance
		Operational consistency	Velocity: completeness, accessibility, efficiency, traceability, availability, recoverability Volume: completeness, accessibility, efficiency, availability, recoverability Variety: accuracy, accessibility, compliance, efficiency, precision, traceability, availability,