



UNIVERSIDAD DE CHILE  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

MODELAMIENTO E IMPLEMENTACION DE UN PLAN DE GOBIERNO DE  
DATOS EN LA UNIVERSIDAD DE CHILE UTILIZANDO EL FRAMEWORK  
DAMA

MEMORIA PARA OPTAR AL TITULO DE INGENIERO CIVIL INDUSTRIAL

MARCO ANDRÉS OLGUÍN DECENCIER

PROFESORA GUIA:  
CAROLINA SEGOVIA RIQUELME

MIEMBROS DE LA COMISION:  
PABLO MARÍN VICUÑA  
SERGIO CELIS GUZMÁN

SANTIAGO DE CHILE  
2023

RESUMEN DE LA MEMORIA PARA OPTAR AL  
TÍTULO DE: Ingeniero Civil Industrial  
POR: Marco Andrés Olguín Decencier  
FECHA: 2023  
PROFESORA GUÍA: Carolina Segovia Riquelme

## **MODELAMIENTO E IMPLEMENTACION DE UN PLAN DE GOBIERNO DE DATOS EN LA UNIVERSIDAD DE CHILE UTILIZANDO EL FRAMEWORK DAMA**

Las prácticas asociadas a la gestión y administración de datos han ido evolucionando en función de los años, siendo un factor determinante los respectivos avances tecnológicos que se han desarrollado en la sociedad. De esta manera, las organizaciones han buscado nuevas formas y metodologías para gestionar la información que poseen, donde el gobierno de datos surge como solución a este desafío, estableciendo los lineamientos para mejorar la gestión de la disponibilidad, integridad, usabilidad y calidad de los datos utilizados.

La Universidad de Chile no es una excepción sobre la gestión de datos, existiendo necesidades específicas con respecto a la calidad de los datos, y el levantamiento de metadata y arquitectura de los datos. En función de estas necesidades, se modeló e implementó un plan de gobierno de datos en la Universidad de Chile, con foco en *data quality*, abordando tanto levantamiento de información crucial para entender la estructura de los atributos a gobernar, como la implementación de herramientas tecnológicas que permitan monitorizarlos y posteriormente repararlos. Asimismo, en cuanto a los alcances pactados para la presente memoria, se aborda un conjunto acotado de datos para ser efectivamente gobernados, los que corresponden a nombres y apellidos, dentro de las tres tablas datos que más transaccionan estos atributos, en específico, las tablas “INDIVIDUO”, “DOTACION” y “DATOS PERSONALES”.

El plan de gobernanza proyectado fue estructurado en dos etapas, la primera de ellas enfocada a poder brindar lineamientos organizativos en cuanto a la estructuración de comités que toman y aprueban decisiones sobre los datos, y la segunda en la implementación técnica, considerando todas las soluciones levantadas para dar cumplimiento al plan. Por un lado, el desarrollo de la implementación técnica brindó una gran cantidad de herramientas asociadas al levantamiento de metadata y arquitectura de datos, donde destacan las matrices CRUD y RACI, mapas de flujo sobre aplicaciones, inventario de activos de datos, entre otros. En cuanto a gestión de calidad del dato, se levantaron formatos transversales de utilización para los atributos a gobernar, además de indicadores de calidad que permitan conocer el actual estado de la calidad de los datos sobre los atributos y tablas a gobernar, para así posteriormente implementarlos en un panel de calidad montado en PowerBI, y finalmente generar acciones reparativas en función de los resultados obtenidos para los indicadores de calidad más críticos.

El panel de calidad levantó la actual situación sobre los atributos y tablas a gobernar, identificando tres indicadores sin cumplimiento de requisitos mínimos de calidad, correspondiendo a exactitud, consistencia y validez, donde en promedio, y considerando un análisis por identificador (RUT), se obtuvieron resultados de cumplimiento del 33%, 57% y 63% respectivamente. En función de estos indicadores críticos, se implementaron soluciones tecnológicas levantadas dentro del transcurso del plan, destinadas a generar limpieza dentro de los atributos y tablas a gobernar, obteniendo mejoras en promedio de 22 puntos porcentuales para exactitud, 21 puntos para consistencia, y 37 puntos para validez, además de brindar estrategias de aseguramiento para abordar los nuevos datos entrantes a las tablas a gobernar.

# Agradecimientos

En primer lugar, quiero agradecer a mis padres Odette y Marco, los que han sido un pilar importante dentro de todo este proceso universitario, y sin quienes hoy no estaría aquí. Les doy muchas gracias por siempre brindarme todo, y apoyarme cuando más lo he necesitado. También quiero agradecer a mi tata, por entregarme su amor, protección y valores, sé que en algún lugar del cielo estará viendo orgulloso este momento.

Quiero agradecer a mi pareja Génesis, la que me ha acompañado todo este proceso, brindándome apoyo y amor cuando más lo necesité, y siendo partícipe de los mejores momentos de mi vida. Gracias por todos los momentos, conversaciones y vivencias que nos han permitido ir evolucionando a la mejor versión de nosotros mismos.

Doy gracias a mis amigos de la universidad, los que han sido un pilar importante dentro de todo este proceso. Gracias por las conversaciones, los podcasts, los buenos momentos, y por compartir la infaltable dosis de fútbol.

Quiero agradecer profundamente a Lucía Moreno, y a todo el gran equipo que compone la Dirección de Datos de la Universidad de Chile, brindándome la mejor primera experiencia laboral que podría haber tenido. Doy gracias por siempre estar ahí para explicarme, enseñarme, y apoyarme cuando lo necesité. Sin ustedes esta memoria no sería posible.

Finalmente quiero agradecer a mis profesores Carolina Segovia y Pablo Marín por permitirme integrar la sección con este tema innovador, y además estar siempre ahí para poder resolver todas mis dudas. Doy gracias por su tiempo, orientación y consejos, los que permitieron lograr la mejor versión de mí en esta última etapa como universitario.

# Tabla de contenido

1	INTRODUCCIÓN .....	1
2	ANTECEDENTES EN LA UNIVERSIDAD DE CHILE .....	2
2.1	Marco institucional .....	3
2.1.1	Contexto tecnológico Universidad de Chile .....	4
3	JUSTIFICACIÓN .....	5
3.1	Encuesta de levantamiento de problemas .....	5
3.1.1	Conocimiento sobre flujos y procesos de datos.....	8
3.1.2	Calidad de datos en bases de datos.....	8
3.2	Propuesta sobre problemáticas levantadas .....	11
4	OBJETIVOS .....	12
5	MARCO CONCEPTUAL.....	12
5.1	Terminología teórica.....	13
5.1.1	Datos .....	13
5.1.2	Metadatos .....	13
5.1.3	Calidad de datos .....	13
5.1.4	DAMA-DMBOK.....	16
5.1.5	Gobierno de datos .....	16
5.2	Terminología técnica .....	17
5.2.1	Bases de datos .....	17
5.2.2	Bases de datos transaccionales .....	17
5.2.3	ETL'S .....	17
5.2.4	Procedimiento almacenado .....	18
5.2.5	API .....	18
5.2.6	Inventario de activos.....	18
5.2.7	Matriz CRUD .....	19
5.2.8	Lenguaje SQL.....	19
5.2.9	SQL Server.....	19
5.2.10	Bases Oracle .....	20
5.2.11	JOB .....	20
5.2.12	Power BI.....	20
5.2.13	DBLINK .....	20
5.2.14	Data Warehouse.....	20
5.2.15	Visual Studio.....	21
5.2.16	Matriz RACI .....	21
5.2.17	Trigger .....	21
5.2.18	JavaScript .....	21
5.2.19	Node.JS .....	22
5.2.20	Formato JSON.....	22
5.3	Frameworks teóricos de apoyo.....	22
5.3.1	DAMA .....	22
5.3.2	Guía técnica MINTIC.....	25
6	ALCANCES .....	27
7	METODOLOGIA .....	27
7.1	Etapa 1: Conformación comités y mesas de trabajo .....	28

7.2	Etapa 2: Implementación técnica gobernanza .....	29
7.2.1	<b>Paso 1: Identificación de datos maestros y/o importantes, definición de metadatos funcionales..</b>	<b>29</b>
7.2.2	<b>Paso 2: Revisión de tratamiento de datos a gobernar dentro de flujos de la universidad.....</b>	<b>30</b>
7.2.3	<b>Paso 3: Perfilamiento .....</b>	<b>30</b>
7.2.4	<b>Paso 4: Despliegue .....</b>	<b>31</b>
8	<b>IMPLEMENTACION.....</b>	<b>32</b>
8.1	Etapa 1: Definición estructura y funciones comités y mesas de trabajo.....	32
8.1.1	<b>Comité estratégico de Gobierno de Datos .....</b>	<b>32</b>
8.1.2	<b>Comité táctico de Gobierno de Datos.....</b>	<b>33</b>
8.1.3	<b>Mesas de trabajo de Gobierno de Datos .....</b>	<b>33</b>
8.2	Etapa 2: (Paso 1) Identificación de datos maestros y/o importantes, definición de metadatos funcionales.....	34
8.2.1	<b>Entrada: Identificación datos a gobernar y mesas de trabajo .....</b>	<b>34</b>
8.2.2	<b>Entrada: Inventario de Activos .....</b>	<b>35</b>
8.2.3	<b>Salida: Diccionario de metadatos .....</b>	<b>36</b>
8.2.4	<b>Salida: Matriz de responsabilidades .....</b>	<b>37</b>
8.3	Etapa 2: (Paso 2) Revisión de tratamiento de datos a gobernar dentro de flujos de la universidad.....	38
8.3.1	<b>Bases de datos institucionales.....</b>	<b>39</b>
8.3.2	<b>Ingreso de data en aplicaciones .....</b>	<b>41</b>
8.3.3	<b>Salida: Flujos de datos sobre aplicaciones .....</b>	<b>41</b>
8.3.4	<b>Salida: Matriz CRUD .....</b>	<b>43</b>
8.4	Etapa 2: (Paso 3) Perfilamiento .....	45
8.4.1	<b>Salida: Levantamiento y definición de indicadores.....</b>	<b>45</b>
8.4.2	<b>Salida: Estandarización de datos .....</b>	<b>46</b>
8.5	Etapa 2: (Paso 4) Despliegue .....	47
8.5.1	<b>Salida: Creación <i>Dashboard</i> de calidad (Monitorización) .....</b>	<b>48</b>
8.5.2	<b>Resultados Monitorización.....</b>	<b>51</b>
8.5.3	<b>Resumen resultados obtenidos .....</b>	<b>60</b>
8.5.4	<b>Análisis resultados monitorización.....</b>	<b>62</b>
8.5.5	<b>Salida: Limpieza de datos .....</b>	<b>63</b>
9	<b>CONCLUSIONES.....</b>	<b>73</b>
10	<b>TRABAJO FUTURO.....</b>	<b>76</b>
11	<b>BIBLIOGRAFÍA .....</b>	<b>78</b>
	<b>ANEXOS.....</b>	<b>82</b>

# 1 INTRODUCCIÓN

La revolución digital ha planteado un cambio transversal en la concepción de cómo llevamos la gestión y la administración dentro de empresas y/o instituciones, y tal como enuncia la asociación internacional de Data Management (DAMA), permeándonos cada vez más al entendimiento de un nuevo paradigma sobre como recolectamos, gestionamos, y finalmente consumimos los datos. Es así como según DAMA (2007), los datos deben ser considerados como un activo fundamental dentro de empresas y/o instituciones, dado que permiten poder alcanzar los objetivos pactados por la organización, y al trabajar minuciosamente su gestión, ofrecerá una toma de decisiones más informada y eficaz. Por otro lado, Britton (2020) destaca que las instituciones necesitan tratar los datos como un activo institucional fundamental dentro de la organización, y que su correcto tratamiento en términos de usabilidad y gestión implicaría una toma de decisiones no sesgada. Se destaca así que los datos surgen como un activo relevante, siendo un factor determinante para crear una ventaja estratégica y competitiva para las empresas, o una mejora en los procesos y la respectiva gestión. Es así como este nuevo paradigma sobre el tratamiento de datos da paso a nuevos métodos y prácticas para entender y gestionar la respectiva información que las organizaciones poseen.

Una de las metodologías utilizadas para gestionar toda esta información que se posee es el gobierno de datos, el que según DAMA (2007), plantea el ejercicio de autoridad y control en función de la gestión de datos, proporcionando una visión integral y transversal de la organización, permitiendo a los responsables de las decisiones dentro de las organizaciones apoyar esta resolución en función de una visión integrada sobre los datos. La gobernanza de datos ha tenido experiencias de éxito en la implementación en organizaciones como la Universidad Tecnológica de Curtin (2), Hospital de la Universidad de Kansas (1), Hospital infantil de Texas (3), entre muchas otras, entregando importantes beneficios en la gestión y administración de datos. Cabe destacar que los programas de Gobierno de datos según DAMA (2007) son únicos, considerando factores asociados a organización y cultura, desarrollo e infraestructura tecnológica, además de las respectivas necesidades que posee cada organización.

La Universidad de Chile no es ajena a estos desafíos que se proponen para las organizaciones, por lo que de esta manera surge la gobernanza de datos como una solución que permite abordar sus respectivas necesidades. De esta manera, para DAMA (2007), existen 10 dimensiones que abarcan la gobernanza, donde para efectos de esta memoria, y en función de los problemas y necesidades de la universidad levantados a lo largo del desarrollo de esta memoria, se dará énfasis en las que están alineadas directamente con estas problemáticas.

Una de las aristas que compone el gobierno de datos, que es particularmente una problemática dentro de la universidad, es la asociada a la gestión de la calidad de los datos, la que según DAMA (2007) corresponde a procesos de apoyo fundamental para el cambio organizacional que propone la gobernanza de datos. Por otro lado, DAMA (2007) también destaca que la calidad de los datos está directamente relacionada con las necesidades que poseen las organizaciones en este aspecto, donde las respectivas acciones asociadas a la gestión, y en particular la medición, monitorización y control debe estar destinadas en función de la identificación de necesidades que estas posean.

Por otro lado, otra de las aristas que comprenden la gobernanza de datos según DAMA (2007), y que particularmente dentro de la universidad no ha existido un desarrollo de esta, corresponde a

gestión de metadata y arquitectura del dato, siendo la documentación de estas dentro de las empresas y/o instituciones las que juega un papel fundamental en la gestión y administración, para así finalmente democratizar conocimiento transversal sobre flujos y procesos a las demás áreas que las componen. Es así como, en primer lugar, Rafael Conde (2018) destaca la arquitectura de datos como un eje importante para brindar ventaja a las organizaciones, la que consiste en poder definir la estructura que tienen los datos, abarcando aristas con respecto a la identificación y esquematización de flujos de datos dentro de las aplicaciones que se poseen para el ingreso de dato. Asimismo, destaca finalmente la gestión de metadatos como un elemento crucial para las organizaciones, proveyendo entendimiento organizacional de estos, colectando e integrando metadatos desde distintas fuentes de datos, para así finalmente disponibilizarlos en función de los atributos que se gobernarán.

Es así como la Universidad de Chile, institución educacional a abordar en la presente memoria, actualmente no presenta levantamiento de ningún plan gobernanza, y en particular, la inexistencia de estrategias que aborden la gestión hacia la calidad del dato, arquitectura de datos, y metadata. De esta manera, se considera que esta institución no está exenta a este gran desafío que se plantea para las organizaciones en la actualidad, de forma que mediante la presente memoria se modelará e implementará un proyecto de gobernanza de datos, tomando aristas como la metadata y arquitectura de los datos, y con un especial énfasis en la calidad de dato, definiendo como alcance la gobernabilidad de algunos atributos pertenecientes a datos personales, y en específico, los asociados a nombres y apellidos, dando así un paso con respecto a este nuevo cambio de paradigma dentro de la institución, buscando nuevas formas de gestionar este activo transcendental, y finalmente democratizar la información dentro de los ejes que conforman la misma.

La estrategia para poder abordar este plan de gobernanza de datos dentro de la Universidad de Chile parte con la definición de los lineamientos para la conformación de comités que propongan y tomen decisiones dentro del plan, para luego dar inicio a la identificación y definición de los elementos potencialmente gobernables, y posteriormente generar un estudio del estado en el que actualmente se encuentran estos datos, para luego definir estándares e indicadores de calidad asociados al dato a gobernar, y finalizar con la implementación de las soluciones tecnológicas pactadas para abordar los problemas levantados.

## 2 ANTECEDENTES EN LA UNIVERSIDAD DE CHILE

La Universidad de Chile es una institución pública de educación superior nacional, en la que existen funciones de docencia, creación y extensión, con un especial énfasis en la investigación y el postgrado. La Universidad de Chile fue la primera universidad creada en nuestro país, fundada en el siglo XIX, específicamente en el año 1842 fue fundada la Universidad de Chile, donde recibe la función de superintendencia de todos los niveles de enseñanza dentro del país, además de darle la misión de propagar la afición para estudios superiores, para así finalmente poder fomentar la investigación y la divulgación en Chile. Para el 1931 se dicta otra ley orgánica, la que finalmente consagra la función docente y científica de la misma, donde desde ese momento se empezaron a crear nuevas facultades, institutos y centros de formación.

Actualmente la Universidad de Chile se rige por el estatuto DFL N°153 de 1981, la que tuvo modificación el 10 de marzo de 2006, dando paso a una nueva institucionalidad, donde desde el punto de vista organizacional los organismos superiores se caracterizan de la siguiente manera:

- Rectoría: Es la máxima autoridad de la universidad, además de su respectivo representante legal, el cual preside el Consejo Universitario y Senado Universitario. El rector está encargado de dictar reglamentos, decretos, y resoluciones; conferir las distinciones, calidades honoríficas, grados y títulos; fijar los aranceles y nombrar al personal académico y administrativo, entre otras atribuciones. (4)
- Prorrector: Es el encargado de asesorar al rector en aspectos económicos, administrativos, jurídicos y estudiantiles. (4)
- Consejo universitario: Es el organismo que tiene por función discutir y aprobar las decisiones del más alto nivel. Está integrado por el Rector, el Prorrector, los Decanos y dos representantes del presidente de la República. (4)
- Senado Universitario: Es un órgano colegiado, encargado de ejercer la función normativa en la Universidad. Este organismo es presidido por el Rector y tiene representación triestamental. Cuenta con 36 miembros, 27 de los cuales son académicos, 7 estudiantes y 2 representantes del personal de colaboración. (4)
- Consejo de Evaluación: Se encarga de coordinar y liderar los procesos de evaluación, calificación y acreditación a nivel institucional e individual, además de constituir comisiones generales y locales. (4)

## 2.1 Marco institucional

La Universidad de Chile actualmente se encuentra dentro del espectro de las entidades educacionales de educación superior dentro del país, específicamente en la agrupación de entidades públicas, manteniendo un constante compromiso con los procesos de enseñanza, creación y divulgación científica a nivel país.

Los actores principales dentro de la universidad son los distintos estamentos que conforman la misma, donde se destacan los estudiantes, funcionarios, y académicos. Los estudiantes son los que actualmente cursan carreras universitarias de pregrado y postgrado dentro de los programas educacionales ofrecidos por la Universidad de Chile, los que además poseen representación en el senado universitario. Por otro lado, los académicos son los encargados de tanto la labor educativa como investigativa de la universidad, realizando docencia, investigación, creación o extensión, integrados a los programas de trabajo de las unidades académicas de la Universidad. Por último, se destaca a los funcionarios, quienes desempeñan las labores de apoyo directivo, profesional, técnico, administrativo y auxiliar. Estos estamentos cumplen con proyectar los objetivos de la Universidad,

y además velar por sus respectivos derechos y necesidades, teniendo cada uno de ellos representación y voz en la toma de decisiones a más alto nivel dentro de la Universidad de Chile.

La instauración de un gobierno de datos en la Universidad de Chile representa un desafío debido a su gran tamaño y complejidad como organización, que se ve acentuada por su alto nivel de descentralización, donde las unidades que componen la misma tienen una gran libertad de acción, y, por tanto, gran cantidad de autonomía en la definición de tecnologías, procesos y normativas sobre sus respectivos datos. Es así como la tarea de centralizar y estandarizar los datos en una institución tan grande y diversa requiere una estrategia cuidadosamente diseñada. Además, se debe considerar la necesidad de involucrar a todas las unidades y departamentos en el proceso de implementación, así como garantizar la interoperabilidad entre sistemas.

### 2.1.1 Contexto tecnológico Universidad de Chile

La universidad ha entrado en un proceso de digitalización importante, reorganizándose jerárquica y administrativamente en función de hacer posible este cambio. Es así como el año 2020 surgió la vicerrectoría de tecnologías e información, la que tiene por objetivo central entregar herramientas tecnológicas y poder desplegar las respectivas soluciones que permitan potenciar el trabajo de la Universidad de Chile en todas las áreas que la componen, buscando así digitalizarla, estableciendo los lineamientos para congregarse una mejor concepción de cambio; fomentando la incorporación de herramientas digitales en todos los procesos internos, contribuyendo así a la eficiencia en la gestión y administración, además de propiciar procesos de aprendizaje que aprovechen la riqueza de las plataformas actuales. (4)

Su trabajo específico se concentra en fomentar la incorporación de tecnologías que favorezcan el funcionamiento coordinado, transversal e integrado de la Universidad; entregando la infraestructura tecnológica necesaria que apoye los procesos inherentes a toda la comunidad universitaria; también promoviendo el acceso y la utilización de los diferentes servicios de tecnologías que permitan enriquecer el trabajo diario de alumnos, académicos y funcionarios. (4)

La vicerrectoría de tecnologías e información se compone de las siguientes direcciones:

- Dirección de Datos: Es la que tiene por tarea recolectar y disponibilizar los datos que produce y recopila la Universidad, generando herramientas que faciliten la toma de decisiones al interior de la Institución, contribuyendo a la construcción de parámetros de alta calidad en los ámbitos de docencia, investigación, creación artística y extensión. (4)
- Dirección de Proyectos Tecnológicos: Es la encargada de diseñar, gestionar e implementar la incubación de proyectos tecnológicos, en conjunto con otros organismos universitarios o unidades académicas cuando sea necesario, a fin de entregar soluciones innovadoras a los requerimientos y problemáticas del quehacer de la Universidad en el cumplimiento de su misión. (4)
- Dirección de Tecnología: Su objetivo es proporcionar soluciones y servicios de tecnología de información y comunicación a la Universidad, manteniendo y operando la

infraestructura, diseñando sistemas que sean eficaces y eficientes de conformidad a la misión de la Institución y proponiendo estrategias, políticas, normativas y estándares de funcionamiento. (4)

### 3 JUSTIFICACIÓN

En función de entender las problemáticas que hoy en día desafían a la Universidad de Chile en relación a la gestión y administración de los datos, se realiza una encuesta de levantamiento de problemas dirigida a los funcionarios de la Vicerrectoría de tecnologías e información de la Universidad de Chile (VTI), los que son actores claves en los procesos y procedimientos tecnológicos sobre la *data*, para así obtener su respectiva percepción sobre los problemas asociados a la gestión y administración dentro de la universidad. Luego de identificar mediante la encuesta los problemas más importantes dentro de la gestión y administración de los datos, se procede a generar un análisis más detallado de estas problemáticas, para finalmente, proponer el gobierno de datos como la respectiva solución para abordarlas, además de complementar con evidencia empírica sobre casos de éxito que implementaron esta solución.

#### 3.1 Encuesta de levantamiento de problemas

Acorde a lo ya mencionado, se desarrolla una encuesta con preguntas relacionadas a identificar las principales dificultades que tienen los funcionarios a la hora de gestionar y administrar datos dentro de la Universidad de Chile. De esta manera, el objetivo de esta encuesta es la de levantar las principales dificultades dentro de la gestión y administración, y así poder priorizar soluciones en función del abordaje de estas problemáticas. De forma referencial, la respectiva encuesta se despliega en la figura 1.



Figura 1 Encuesta funcionarios VTI

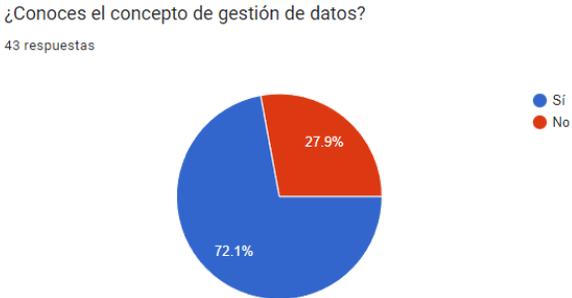
Esta encuesta contó con 4 preguntas relacionadas a en primer lugar poder levantar el respectivo conocimiento de los funcionarios de la Vicerrectoría de tecnologías de la Universidad de Chile (VTI) sobre conceptos asociados a la gestión de datos y el gobierno de datos, y, en segundo lugar,

levantar los principales dolores y/o problemáticas asociadas al trabajo en función de los datos. La cantidad de respuestas obtenidas fue de 43, representando el 57.33% de los funcionarios a los que se envió esta encuesta. Esta encuesta estuvo disponible para ser respondida durante una semana, y en términos de confidencialidad, se consideró que las respuestas enviadas por los funcionarios se mantuvieran anónimas.

Las preguntas que se encontraban en la respectiva encuesta fueron las siguientes:

- ¿Conoces el concepto de gestión de datos?
- ¿Conoces el concepto de gobierno de datos?
- ¿Cuáles crees que son los problemas sobre datos que actualmente le acomplejan como funcionario de la VTI? (Puede seleccionar más de una opción).
- Dentro de la institución ¿Cuál cree que es la dificultad más frecuente en torno a la calidad de los datos existentes? (puede marcar más de uno).

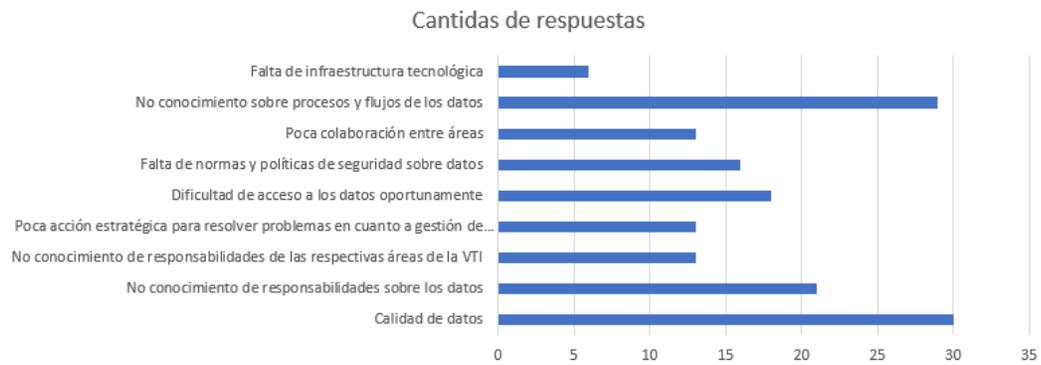
Los resultados de la encuesta levantaron el conocimiento de los respectivos funcionarios sobre los conceptos de gestión y gobernanza de datos, además de las problemáticas más importantes de parte de los funcionarios con respecto a los datos, las cuales se ven reflejadas en las figuras 2, 3, 4 y 5.



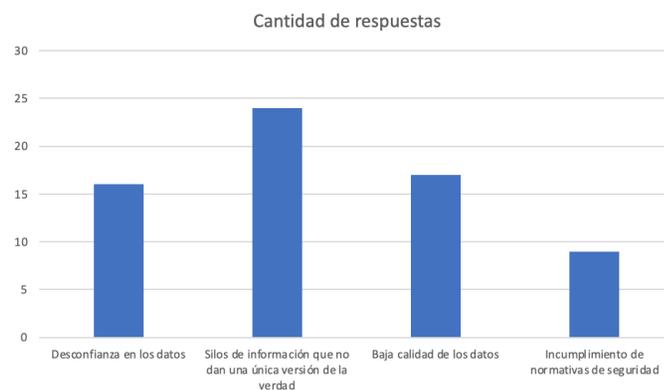
*Figura 2 Encuesta funcionarios VTI*



*Figura 3 Encuesta funcionarios VTI*



*Figura 4* ¿Cuál crees que son los problemas sobre datos que actualmente le acomplejan como funcionario de la VTI?



*Figura 5* Dentro de la institución ¿Cuál cree que es la dificultad más frecuente en torno a la calidad de los datos existentes?

En cuanto a las respuestas recibidas dentro de las preguntas mencionadas en las figuras 2 y 3, es claro que existe un conocimiento importante del concepto de gestión de datos, sin embargo, el concepto de gobierno de datos representa un desconocimiento mayor.

Por otro lado, en relación con las preguntas asociadas a las problemáticas dentro de la gestión y administración de *data*, en la figura 4 los resultados reflejaron una actual convergencia en que la calidad de los datos no cumple sus expectativas, representando un 30% de la elección de problemáticas, además de expresar un desconocimiento sobre los respectivos procesos y flujos implicados en los datos de la universidad, representando un 27%. Por otro lado, en la figura 5, la mayoría de los funcionarios consultados congregan que en general no existe un acuerdo con respecto a la única verdad de los datos, derivando en que existan muchas fuentes de estos que causan errores de consistencia en los datos, además de problemas asociados a la calidad de los datos y la respectiva desconfianza en estos.

Por tanto, se puede apreciar como los problemas de calidad de datos y el no conocimiento de flujos y procesos resultaron ser las problemáticas que más acomplejan hoy en día a los funcionarios de la VTI, y dentro de un marco organizacional y de gestión, se hace indispensable poder brindar una nueva forma sobre cómo se entiende la información, y considerar la importancia de enmarcarla como un activo fundamental. Luego de identificar estas problemáticas con respecto a la gestión y administración de datos levantadas por la encuesta, se propone en las siguientes dos subsecciones

abordar con mayor detalle estas problemáticas (Calidad de datos y el no conocimiento de flujos y procesos), para así sustentar estos hallazgos obtenidos.

### 3.1.1 Conocimiento sobre flujos y procesos de datos

Tal como menciona Rafael Conde del Pozo (2018), dentro de un plan de gobernanza de datos es de gran importancia poder integrar aristas relacionadas a los metadatos de los datos, y su respectiva arquitectura, incluyendo documentación que permita finalmente generar una visión unificada de los flujos y procesos involucrados en la gestión y administración de los datos a gobernar.

Es importante destacar que dentro de la gestión y administración de la Universidad de Chile no existe un levantamiento y posterior gestión de estas aristas y tal como se menciona en la encuesta realizada, uno de los problemas que actualmente afecta la gestión de los datos dentro de la Universidad de Chile, es la no existencia de un levantamiento de los flujos, procesos, y arquitectura de los sistemas tecnológicos de la universidad. De esta manera, no existen:

- Diagramas de flujo de datos para las aplicaciones que actualmente ingresan o modifican datos dentro de las bases de datos
- Información sobre los permisos de accesos de las respectivas cuentas con atribuciones de ingreso, creación, modificación y eliminación de datos sobre las bases de datos transaccionales
- Levantamiento de metadata en función de datos importantes para la gestión y administración.
- Levantamiento de activos de información en función de datos importantes para la gestión y administración.

Es así como la inexistencia de esta documentación asociada a la metadata y arquitectura de los datos implica en que las distintas áreas de negocio de la universidad no conozcan los flujos, procesos, y responsables de las demás áreas sobre los datos, generando así una incomunicación dentro de éstas. De este modo, la demanda y oferta de datos para la gestión se ve afectada debido al no conocimiento sobre dónde se encuentran los datos, y con quién poder consultarlos.

Para generar una intercomunicación más fluida dentro de las distintas áreas de negocio, se propone generar un levantamiento y democratización de esta información, además de poder armonizar los flujos en función de una mejora sustentable para la comprensión de la propia información que como institución se posee.

### 3.1.2 Calidad de datos en bases de datos

A nivel global, la calidad de los datos ha sido una arista que cada vez ha tomado mayor impacto dentro de las organizaciones, debido a los múltiples beneficios asociados a la gestión, administración, y la toma de decisiones dentro de estas mismas. La calidad de los datos según DAMA (DAMA Internacional, 2009) se trata de una etapa esencial para ayudar en la

administración del cambio organizacional de manera crítica. Cuando se cambia la perspectiva de los negocios, se integran estrategias corporativas de negocios, o se realizan fusiones, adquisiciones y asociaciones, es necesario que las funciones de TI unifiquen y fusionen las fuentes de datos, generen copias de información importante y combinen datos retrospectivamente o de manera integral.

Este proceso de consideración en cuanto al valor que posee la calidad de los datos, y los aportes que tendrían dada la gestión de su respectiva calidad han ido evolucionando a lo largo de los años, en función de problemáticas que han surgido dentro de las organizaciones a nivel mundial, las cuales demuestran magnitudes importantes en términos económicos y de gestión. Es así como a modo de ejemplo en el 2002, las compañías de Estados Unidos supusieron pérdidas de alrededor de 611 billones de dólares debido a la baja calidad que tenían en sus bases de datos (5). Asimismo, se destaca que, en el 2011, la baja calidad de los datos en las bases transaccionales de las organizaciones era la principal razón por la que las iniciativas de negocio fracasaban en poder alcanzar los objetivos definidos por ellos mismos. (6) Finalmente, se destaca que, en 2014, el 59% de las empresas apuntaban a la calidad de sus datos como una barrera para adoptar nuevas herramientas tecnológicas asociadas a la inteligencia de negocios. (7)

Es así como la Universidad de Chile no es ajena a este problema que actualmente enfrentan las organizaciones, y tal como se expresa en la encuesta levantada, la noción que tienen los funcionarios es que existe una problemática con respecto a la calidad de los datos. Para tener una referencia de entrada a la actual calidad de los datos en la Universidad de Chile, se desarrolló un análisis preliminar de calidad considerando 3 indicadores de calidad, y uno de los atributos a gobernar, en específico, el atributo “NOMBRE”, escogido para esta referencia preliminar en función de la importancia y transversalidad que tiene este atributo para una gran cantidad de procesos y flujos dentro de la universidad, el que corresponde a literalmente capturar el nombre de las personas que se encuentran dentro de las bases transaccionales. Estos indicadores fueron extraídos y definidos del *framework* DAMA, para luego mediante queries realizadas en lenguaje SQL, poder generar el cálculo de estos respectivos indicadores dentro de una de las tablas más pobladas con registros, en particular, la tabla “INDIVIDUO”, la que en secciones posteriores de la memoria se introducirá con mayor precisión. Los indicadores seleccionados para este análisis preliminar fueron considerados de la siguiente manera:

- **Compleitud:** Indica el porcentaje de registros que poseen un valor no nulo dentro de la tabla.

$$\text{Compleitud} = \frac{\sum \text{Valores no nulos}}{\text{Total registros}} \quad (4.1)$$

- **Validez:** Indica el porcentaje de registros que poseen el formato definido para el respectivo análisis sobre las tablas.

$$\text{Validez} = \frac{\sum \text{Valores que cumplen formato}}{\text{Total registros}} \quad (4.2)$$

- Consistencia: Indica el porcentaje de registros que cruzan correctamente con las demás tablas para el atributo “Nombre”, dentro de la misma base de datos.

$$\text{Consistencia} = \frac{\sum \text{Valores que cruzan correctamente con otras tablas}}{\text{Total registros que cruzan}} \quad (4.3)$$

Para el cálculo de estos indicadores en función del atributo “NOMBRE”, se obtuvieron los siguientes resultados expuestos en las figuras 6 y 7.



Figura 6 y 7 Indicadores validez, completitud y consistencia

Es así como cuantitativamente se puede apreciar resultados con márgenes de error con respecto a estos 3 indicadores de calidad, y en particular para los indicadores de validez y consistencia, y que sumado a los resultados obtenidos por la encuesta de levantamiento de problemas, plantea una situación que potencialmente podría estar replicándose en otros conjuntos de datos con respecto a la calidad de estos dentro de las bases de datos de la universidad.

Para dar con el origen de esta problemática asociada a la calidad de los datos, hay que recurrir a entender cómo ingresan los respectivos datos a las bases transaccionales de la institución. Actualmente dentro de la Universidad de Chile existe una gran cantidad de sistemas tecnológicos legados del pasado, y en particular, aplicaciones que tienen permisos de ingreso y/o modificación de datos, las que fueron desarrolladas en diferentes contextos tecnológicos, heredando así aplicaciones que no se adecuan a los estándares actuales en términos de calidad. Por otro lado, estas respectivas aplicaciones no han pasado por procesos de actualización (Modificación de aplicaciones) que reflejen las respectivas necesidades que se demandan hoy en día, lo que ha incurrido en un problema con respecto a la recolección de datos. De esta manera, muchas aplicaciones que actualmente se encuentran activas dejan a libertad del usuario el ingreso de datos en los formularios que proveen estas aplicaciones, provocando que una gran cantidad de estos datos no posean formato ni validez mínima para una toma de decisiones dentro de proyectos y programas de la Universidad de Chile.

Asimismo, cabe destacar que la evolución tecnológica no ha sido homogénea en la universidad, existiendo aún muchas áreas donde existe una precarización digital, lo que se traduce

en que todavía se sigan generando prácticas de llenado manual de escrito a digital, lo que eventualmente podría desencadenar errores que afecten la calidad del dato.

Estos problemas asociados a la calidad de los datos que actualmente se tienen en las bases transaccionales de la Universidad de Chile, se reflejan en la no existencia de herramientas de monitorización para medir y desplegar informes sobre la calidad de los respectivos datos, ni en la realización de acciones de limpieza masivas dentro de las respectivas tablas que almacenan los datos.

## 3.2 Propuesta sobre problemáticas levantadas

Es así como las necesidades identificadas por el levantamiento realizado en la encuesta dirigida a los funcionarios de la VTI, y tras su respectivo análisis en detalle, expone cómo en general la calidad de los datos y el desconocimiento sobre flujos y procesos sobre estos se presentan como una problemática dentro de la gestión y administración dentro de la universidad, donde en términos de calidad, se pierde la fiabilidad por los datos, provocando que todas las áreas de la Universidad de Chile que consumen estos datos destinen tiempo en limpiar superficialmente estos, o no puedan usarlos, y por otro lado, con respecto a flujos y procesos de los datos, al no conocer su funcionamiento y responsables, generan una incomunicación entre las áreas que tienen injerencia sobre los datos.

De esta manera, la gobernanza de datos enfocada en particular a la gestión de calidad de los datos, y tocando aristas asociadas a la gestión de metadata y la gestión de la arquitectura de los datos, presentan una solución a estas necesidades, donde en primer lugar se pueda abordar los problemas de calidad enunciados, considerando estrategias de monitorización y de limpieza, tanto a corto como a largo plazo, sopesando problemas ya sea de manera superficial como de raíz, y en segundo lugar, abordar la documentación de una gran cantidad de información determinante para mejorar el entendimiento de la arquitectura y metadata de los datos, y de esta manera, abordando el levantamiento y democratización de los flujos y procesos asociados a los datos a gobernar.

Para demostrar la eficacia de la gobernanza de datos como una solución factible para lograr una gestión de datos más eficaz, se ejemplifican 2 casos de éxito. El primer caso es el del Hospital de la Universidad de Kansas, el cual enfrentó problemáticas con respecto a la cantidad de tiempo necesaria para recopilar, adquirir y validar manualmente datos de distintas fuentes, además de las respectivas problemáticas asociadas a la calidad del dato, generando una desconfianza transversal en la fiabilidad de los datos, y de esta manera, se modeló e implementó un plan de gobernanza de datos, que permitió gestionar y mejorar la calidad de datos transversalmente dentro del hospital, además de levantar información crítica para apoyar la gestión (1). El segundo caso de éxito es el del Hospital Infantil de Texas, organización que tuvo problemas con respecto a la democratización de datos e información, limitando la capacidad para poder realizar análisis predictivos sobre la *data*, y debido a esta razón, se propuso la implementación de un plan de gobernanza de datos, el que en términos de resultados, implicó en una mejora del 25% con respecto al número de accesos de datos en un año, además de una mejora de un 81% en el tiempo de respuestas para nuevas solicitudes de informes de datos. (3)

Dada la efectiva implementación en estos respectivos casos expuestos, se propone el modelamiento y posterior implementación de un plan de gobernanza de datos en función de las necesidades

levantadas mediante la encuesta realizada a los funcionarios de la VTI, y de esta manera, abordando en particular a la calidad de los datos y la gestión de arquitectura y metadatos de estos.

## 4 OBJETIVOS

Objetivo General:

- Modelar e implementar un programa de Gobierno de Datos en la Universidad de Chile utilizando el *framework* DAMA, abordando atributos relacionados a nombres y apellidos, y con foco en calidad del dato.

Objetivos Específicos:

- Establecer lineamientos organizativos de toma y aprobación de decisiones para una correcta implementación del plan de gobierno de datos.
- Priorizar y seleccionar los datos a gobernar en las primeras iteraciones del plan.
- Levantar el tratamiento actual en flujos sobre aplicaciones que involucren a los datos a gobernar.
- Perfilar datos a gobernar, estableciendo estándares e indicadores de calidad asociados a estos.
- Implementar soluciones tecnológicas de monitorización y limpieza para gestionar la calidad de los datos a gobernar.

## 5 MARCO CONCEPTUAL

En el presente capítulo se entenderá a cabalidad el sustento teórico que comprende todo el plan de gobernanza de datos en la Universidad de Chile. Para esto, en primera instancia se presentará todo el glosario terminológico que entiende la proyección de este plan, para después mencionar en el siguiente subcapítulo los distintos *frameworks* disponibles, y el por qué el *framework* DAMA fue utilizado para su respectivo desarrollo e implementación.

## 5.1 Terminología teórica

### 5.1.1 Datos

Según DAMA (DAMA International, 2009) los datos representan simbólicamente hechos en forma de texto, números, gráficos, entre otros. El concepto de datos proviene del plural de la palabra *datum*, de origen latín, que significa “un hecho”, y de ahí la definición de estructuras en función de este. Estos respectivos hechos son capturados, almacenados y finalmente expresados como datos, entendiéndose como la materia prima que utilizan los consumidores de datos. Dentro de este mismo contexto, DAMA entiende que los datos pueden ser inexactos, incompletos, desactualizados o incluso incomprendidos. Es así como la gestión de la data, y en específico la calidad de estos juega un rol realmente fundamental para asegurar una toma de decisiones sin sesgo. Es tal la importancia que tienen hoy en día los datos para las organizaciones, que en su defecto deben ser considerados como activos importantes dentro de estas. Tal como menciona DAMA “Los activos de datos son los recursos con valor reconocido bajo el control de un individuo u organización. Los activos de la empresa permiten alcanzar los objetivos de la empresa y por lo tanto necesitan ser manejados cuidadosamente. La captura y el uso de dichos activos se controlan cuidadosamente y las inversiones en estos activos se aprovechan eficazmente para alcanzar los objetivos empresariales”. (8)

### 5.1.2 Metadatos

Según DAMA, los metadatos son los que ayudan a establecer el contexto de los datos, contribuyendo directamente a mejorar la calidad de la información. La gestión de los activos de información incluye la gestión de datos y sus metadatos. (DAMA Internacional, 2009)

Por otro lado, PowerData los define como “información que permite brindar información sobre los datos, y así poder conocer la estructura de ellos, cómo verificar la validez sintáctica del mismo, etc.” permitiendo así describir otros datos o generar “datos sobre datos”, describiendo una contextualización de la información que se tiene dentro de la organización. (9)

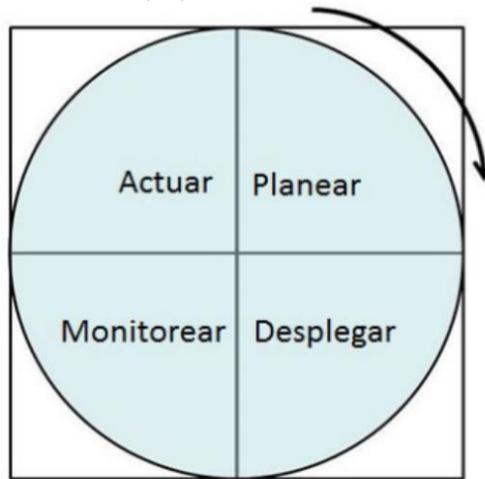
Es así como hoy en día las empresas se centran más en la necesidad y la importancia de los metadatos en sus respectivos procesos y flujos. Por otro lado, también evalúan cómo incorporar los metadatos dentro de las fuentes estructuradas tradicionales, además de las fuentes no estructuradas, para así generar un mejor entendimiento del contexto general.

### 5.1.3 Calidad de datos

Según PowerData (2016) la calidad de datos hace referencia a los procesos, técnicas, algoritmos y operaciones encaminados a mejorar la calidad de un conjunto de información, pudiendo ser una base de datos, un sistema de información o un data warehouse, entre otros. Para que exista una buena calidad en los datos, éstos deben cumplir una serie de criterios definidos, concretamente, indicadores que proyecten la calidad de estos, cuyo propósito final sea poder generar una monitorización de la información que pueda guiar una correcta toma de decisión en el tratamiento de estos. (PowerData, 2016)

Para DAMA, la calidad de datos comprende “un proceso de apoyo crítico en la gestión del cambio organizacional” donde su función radica en “definir los parámetros para especificar los niveles aceptables de calidad en función de indicadores de calidad, para así finalmente satisfacer las necesidades del negocio y garantizar que la calidad de datos cumple con estos niveles”. (10)

Para poder entender una buena gestión de la calidad de los datos dentro de la organización, DAMA detalla los pasos sobre “una versión del ciclo de Deming. Deming, uno de los escritores fundamentales en la gestión de calidad, donde se propone un modelo de resolución de problema conocido como “planificar-hacer-estudiar-actuar” o ”planificar-hacer-verificar-actuar”, que es útil para la gestión de calidad de los datos”. (11) Este ciclo se detalla en la figura 8.



*Figura 8 Ciclo de calidad del dato (DAMA, 2009)*

El ciclo propuesto comienza identificando los actuales problemas de datos que son críticos para el logro de los objetivos de la organización, para así definir las dimensiones de calidad de datos a utilizar en las distintas iteraciones del proceso, además de identificar todas las respectivas reglas del negocio asociadas que permitan una completa gestión de la calidad de los datos de la institución. (DAMA Internacional, 2009). En la etapa de planeación, se estudian los problemas encontrados, evaluando el alcance de estos, además de dimensionar el impacto de estos problemas dentro de los flujos y procesos, para así poder abordar distintas alternativas de solución para la calidad de estos. (DAMA Internacional, 2009). En la fase de despliegue, se realiza el perfilamiento de los datos involucrados dentro del ciclo, levantando herramientas de monitoreo y limpieza. En esta fase el equipo de calidad de datos puede rediseñar algunos procesos que son causa de los errores de datos. Finalmente, cuando no es posible corregir errores en su origen, se debe corregir estos errores en el primer punto después del proceso de origen. (DAMA Internacional, 2009). El monitoreo está relacionado con la supervisión activa de la calidad de los datos dentro de los flujos y procesos comprometidos, pudiendo así verificar activamente los actuales niveles de calidad. En el caso que la calidad de los datos sufran bajas considerables, los administradores de datos deben ser notificados para que se puedan tomar medidas de corrección. (DAMA Internacional, 2009). Finalizando el ciclo la etapa actuar es para tomar medidas para abordar y resolver los nuevos problemas de calidad de datos que emergen luego del ciclo. Las siguientes iteraciones comienzan como nuevos conjuntos de datos que estarán bajo investigación, o como nuevos requerimientos de calidad de datos que son identificados para los conjuntos de datos existentes. (DAMA Internacional, 2009)

Por otro lado, DAMA es enfático en que existe una necesidad urgente de “segmentar las reglas de negocio de acuerdo con las dimensiones de calidad de los datos, implementando indicadores que puedan eventualmente medir la calidad, incluyendo detalles sobre los alcances de la respectiva medición”. Para abordar esta necesidad, DAMA expone los siguientes 11 criterios de calidad: (12)

- Exactitud: Se refiere a como los datos representan correctamente información de la "vida real" que estos modelan. Un ejemplo concreto de este criterio podría ser la comparación de los valores contra una base de datos previamente validada. (DAMA Internacional, 2009)
- Consistencia: Hace referencia a garantizar que los valores de datos en un conjunto de datos son coherentes con los valores en otro conjunto de datos. (DAMA Internacional, 2009)
- Vigencia: Se refiere al grado en el que la información está actualizada al día en el que se usa el respectivo dato. (DAMA Internacional, 2009)
- Precisión: Se refiere al respectivo nivel de detalle del dato. Los datos numéricos pueden necesitar precisión en dígitos significativos. (DAMA Internacional, 2009)
- Privacidad: Hace referencia a la necesidad de control de acceso y de supervisión del uso de los respectivos datos de la organización. (DAMA Internacional, 2009)
- Integridad Referencial: Es cuando todas las referencias o relaciones destinadas a partir de los datos de una columna de una tabla hacia los datos de otra columna de la misma o de diferente tabla sean válidas. (DAMA Internacional, 2009)
- Oportunidad: Es la expectativa de accesibilidad y disponibilidad de la información, es decir, la medición del tiempo transcurrido entre solicitudes y oportuno acceso a estos. (DAMA Internacional, 2009)
- Unicidad: Hace referencia a que ninguna entidad de datos exista más de una vez dentro del mismo conjunto de datos. (DAMA Internacional, 2009)
- Validez: Se refiere a si los datos se presentan en un formato que es consistente con el dominio de los valores, respetando los respectivos estándares definidos para los datos a gobernar. (DAMA Internacional, 2009)

## 5.1.4 DAMA-DMBOK

DAMA se autodefine como “una asociación global sin fines de lucro, independiente de proveedores, de profesionales técnicos y comerciales dedicada a promover los conceptos y prácticas de la gestión de información y datos” donde su principal labor es la de “promover los conceptos y las prácticas de gestión de datos y respaldar a los miembros de DAMA-I y sus organizaciones para abordar sus necesidades”, facilitando así el desarrollo de cuerpos de conocimiento a través de su comunidad de expertos, así como el “desarrollo de programas de certificación y capacitación”. (13)

Por otro lado, en función de poder generar una democratización de los conocimientos que posee esta organización de profesionales del data management, redactaron un completo *framework* guía para la gestión de buenas prácticas acerca de los datos. Es así como el DAMA-DMBOK “define una visión estándar de industria para las funciones de gestión de datos, la terminología y las mejores prácticas, sin detallar métodos y técnicas específicas” donde “si bien DAMA-DMBOK no es una autoridad completa en ningún tema específico, señalará a los lectores otras publicaciones, artículos y sitios web ampliamente reconocidos para obtener más información”. (13)

## 5.1.5 Gobierno de datos

Según Power Data (2016), el gobierno de datos se entiende como la capacidad de una organización para gestionar el conocimiento que tiene sobre su información, de forma que se pueda responder a preguntas acerca de la Metadata de la información y el respectivo origen de esta información.

Por otro lado, DAMA *Framework* nos menciona la gobernanza como “el ejercicio de la autoridad y el control (planificación, el seguimiento y la aplicación) a través de la gestión de los activos de datos. La función de Gobierno de Datos guía el cómo se llevan a cabo todas las demás funciones de gestión de datos. Gobierno de Datos es de alto nivel, la administración ejecutiva de los datos”. (14)

El instituto MDM define Data Governance como “la orquestación estructurada tanto de personas, procesos y tecnologías, las que permiten a una organización aprovechar la data como activo importante de la empresa”. (15)

El Instituto de Gobierno de datos (DGI) afirma que el gobierno de datos “es un sistema de derechos de decisión y rendición de cuentas para los procesos relacionados con la información, ejecutado según acuerdos de modelos que describen quién puede tomar las acciones con que información, cuándo, en qué circunstancias, y con qué métodos”. (16)

Por otro lado, según IBM, “el gobierno de datos ayuda a la gestión de información organizacional durante todo el ciclo de vida, tanto desde su creación como hasta su eliminación. El gobierno de datos aborda los problemas que desafían las iniciativas de gestión de registros, descubrimiento electrónico, conformidad, optimización del almacenamiento y migración de datos”. (17)

Finalmente, según ORACLE el gobierno de datos es la “disciplina que tiene por objetivo hacer cumplir las políticas y procedimientos para el uso y la gestión de los datos, coordinando las personas, los procesos y la tecnología con el fin de posicionar la información como un recurso de

valor que satisfaga las necesidades de información de los interesados y que permita cumplir los objetivos de la organización”. (18)

Se puede afirmar entonces que el gobierno de datos hace referencia a la búsqueda de una gestión eficiente de la información que se posee dentro de la organización, ofreciendo una respuesta a la revolución de los datos, logrando configurar holísticamente una visión general y transversal de la disponibilidad, usabilidad, seguridad e integridad de los datos que posee la organización.

## 5.2 Terminología técnica

Tal como se expresa en capítulos anteriores, el gobierno de datos que se propone dentro de esta memoria converge tanto conceptos teóricos, como técnicos asociados a infraestructura tecnológica. Es así como se hace indispensable poder introducir estos conceptos que eventualmente harán entender al lector la complejidad informática de algunos procesos involucrados.

### 5.2.1 Bases de datos

Según ORACLE, las bases de datos son “una recopilación estructurada de información o datos estructurados, los que se almacenan en un sistema informático” y que por otro lado “suelen utilizar estructuras de filas y columnas en una serie de tablas para potenciar la eficacia del procesamiento de información y la consulta de datos dentro de estas mismas” (19)

Es así como, las bases de datos permiten el almacenamiento estructurado de datos, para así garantizar la integridad de los datos y facilitar las tareas tanto de desarrolladores como de usuarios.

### 5.2.2 Bases de datos transaccionales

Las bases de datos transaccionales según PowerData se caracterizan por “permitir llevar a cabo un gran número de transacciones cortas en línea, haciendo posible un procesamiento de consultas muy rápido, manteniendo la integridad de los datos en entornos de acceso múltiple y garantizando unos niveles de efectividad muy elevados. Una base de datos de este tipo cuenta con información actual y detallada”. (20)

Es así como las bases de datos transaccionales congregan una arista realmente importante dentro de la implementación de este proyecto, esto en función de que la actual arquitectura tecnológica que presenta la Universidad de Chile para resguardar sus respectivos datos es dada por bases de datos transaccionales.

### 5.2.3 ETL'S

Las ETL según PowerData “son una parte de la integración de datos, siendo un elemento importante cuya función completa el resultado de todo el desarrollo de la cohesión de aplicaciones y sistemas.”. La sigla ETL significa; extraer (E), transformar (T), y cargar (L) (21).

Extraer hace referencia a la extracción de datos desde los sistemas de origen, analizando los datos extraídos obteniendo un chequeo, interpretando este chequeo, para así verificar que los datos extraídos cumplen la pauta o estructura que se esperaba, y finalmente convertir estos datos a un formato preparado para iniciar el proceso de transformación. (21)

Por otro lado, la fase de transformación “aplica una serie de reglas de negocio o funciones sobre los datos extraídos para convertirlos en datos que serán cargados”, estableciendo reglas funcionales para poder adecuarse a las distintas fases de destino. (21)

Finalmente, la fase de cargar según PowerData entabla que “los datos procedentes de la fase anterior (fase de transformación) son cargados en el sistema de destino. Dependiendo de los requerimientos de la organización, este proceso puede abarcar una amplia variedad de acciones diferentes”. (21)

Es así como los ETL’s comprenden una parte importante dentro de todos los procesos de calidad implicados en *data*, siendo fundamental para poder congrega datos desde distintas fuentes, implicando estándares y normas de carga fundamentales dentro de los alcances del plan de gobierno de datos proyectado para la Universidad de Chile.

#### 5.2.4 Procedimiento almacenado

Un procedimiento almacenado es “un conjunto de instrucciones que se pueden llamar desde otras consultas o desde otros procedimientos almacenados. Un procedimiento puede tomar argumentos de entrada y mostrar valores como resultados”. (22)

Por otro lado, DAMA destaca que un procedimiento almacenado en “una base de datos puede enviar un mensaje resultante de la ejecución de una consulta, procedimiento almacenado o función en otra base de datos, con los resultados devueltos de forma asíncrona al procedimiento de llamada. Este enfoque es relativamente fácil de implementar, fiable, escalable y funciona bien.” (23)

#### 5.2.5 API

Según RedHat las API son “conjuntos de definiciones y protocolos que se utilizan para diseñar e integrar el software de las aplicaciones. Suele considerarse como el contrato entre el proveedor de información y el usuario, donde se establece el contenido que se necesita por parte del consumidor (la llamada) y el que requiere el productor (la respuesta)”. Es así como las API resultan una opción indispensable para poder acceder información desde un equipo remoto, además de proveer una solución factible ante la demanda de datos. (24)

#### 5.2.6 Inventario de activos

Según el MinTic, los inventarios de activos disponen una herramienta fundamental para poder identificar, clasificar, y finalmente gestionar los activos fundamentales que tienen implicancia dentro de los datos definidos, además de declarar los roles y responsabilidades que tienen las distintas áreas sobre estos activos. Finalmente, este inventario también considera clasificación de riesgo para cada activo identificado, los que se definirán en función de criterios previamente establecidos. (25)

Por otro lado, en cuanto a las definiciones importantes a considerar dentro del inventario de activos, se destaca la asignación de roles y responsabilidades, donde dentro de estas se destacan 3 clasificaciones para términos de esta memoria, las que se desprenden a continuación:

- Custodio del dato: Área encargada de administrar y salvaguardar la seguridad de los datos a gobernar, siguiendo los parámetros definidos por los respectivos propietarios de la información, tales como copias de seguridad, asignación de permisos de acceso a bases de datos, modificación, entre otros. (26)
- Propietario del dato: Área que tiene la responsabilidad de garantizar que la información relacionada con los servicios de procesamiento de información se clasifiquen correctamente, además de definir y revisar periódicamente las restricciones y clasificaciones del acceso dentro de bases de datos. (26)
- Responsable del activo: Es el área encargada de hacer efectivas las restricciones y clasificaciones de acceso a bases de datos que fueron previamente definidos por el propietario del activo y/o dato. (26)

### 5.2.7 Matriz CRUD

Según MinTic, la matriz CRUD identifica “qué tratamiento realiza cada cuenta sobre las bases de datos de la institución, utilizando operaciones dentro de los permisos como: crear, leer, actualizar y eliminar”. (27)

Por otro lado, según DAMA, la matriz CRUD es la que explicita las “áreas del negocio que tienen la responsabilidad de crear, actualizar, eliminar y usar datos dentro de tablas específicas”. Es así que, con esta herramienta, se puede obtener un panorama general sobre como las cuentas de usuarios dentro de las bases de datos accionan dentro de estas. (28)

### 5.2.8 Lenguaje SQL

Según Microsoft, SQL es “un lenguaje de computación para trabajar con grandes conjuntos de datos y las respectivas relaciones entre ellos” siendo un “estándar internacional reconocido por organismos de estándares como ISO y ANSI”. (29)

Por otro lado, según IBM es un “lenguaje estandarizado que tiene por función definir y manipular los datos de una base de datos relacional. La base de datos se crea como un conjunto de tablas, las relaciones se representan mediante valores en las tablas y los datos se recuperan especificando una tabla de resultados que puede derivarse de una o más tablas base”. (30)

### 5.2.9 SQL Server

Una de las herramientas fundamentales para el desarrollo de esta memoria es SQL Server, el que “es la alternativa que ofrece Microsoft como un sistema gestor de bases de datos”, donde, en específico, es un “sistema de gestión de base de datos relacional desarrollado a modo servidor, que

tiene por función dar servicio a otras aplicaciones de software que pueden funcionar, ya sea en el mismo ordenador o en otro computador remoto a través de una red”. (31)

### 5.2.10 Bases Oracle

Oracle *Database* es un “sistema que permite gestionar bases de datos relacionales de Oracle. Como software de bases de datos”, donde además “permite brindar optimización en la seguridad y administración de bases de datos, creando así esquemas lógicos y estructurados a los que solo pueden acceder usuarios previamente autorizados”. (32)

### 5.2.11 JOB

Un JOB dentro del contexto SQL es un “proceso que se ejecuta automáticamente en una programación establecida bajo control del sistema gestor de bases de datos. Por medio del uso de la programación de este trabajo automático, se puede determinar cuándo y en qué orden se realizarán las tareas administrativas”. Por otro lado, las “tareas que se programan suelen ser tareas que toman mucho tiempo en su ejecución, y por tanto acostumbran a programarse por la noche cuando los recursos de uso de las bases de datos están más vacíos”. (33)

### 5.2.12 Power BI

Según Microsoft, la propia compañía que ha desarrollado Power BI, explicita que es un “conjunto de servicios de software, aplicaciones y conectores que funcionan conjuntamente para convertir orígenes de datos sin relación entre sí en información coherente, interactiva y atractiva visualmente”, donde además destacan que “sus datos pueden ser una hoja de cálculo de Excel o una colección de almacenes de datos híbridos locales y basados en la nube”. Por otro lado, “Power BI permite conectarse con facilidad a los orígenes de datos, visualizar y descubrir qué es importante y compartirlo con cualquiera o con todos los usuarios que desee.”. (34)

### 5.2.13 DBLINK

Un DBLink es un “tipo de objeto que ajusta y genera una conexión desde una base de datos a una externa a esta”, donde el principal propósito es “poder ocultar el detalle de los ajustes de conexión necesarios, ofreciendo ingreso a todos los recursos disponibles en otras bases de datos, independientemente del servidor en que se encuentren.”. (35)

### 5.2.14 Data Warehouse

Según Power Data, un data warehouse es “un almacén de datos donde generalmente una organización mantiene una gran cantidad de información”. Por otro lado, también destaca que “los datos almacenados dentro de un data warehouse deben alojarse de forma segura, fiable, fácil de recuperar y fácil de administrar.”. (36)

Por otro lado, según DAMA, un data warehouse es “una combinación de dos componentes principales. El primero hace referencia a una base de datos integrada para dar soporte continuo a la toma de decisiones dentro de la organización y/o empresa. El segundo está relacionado con los

programas usados para recolectar, limpiar, transformar y almacenar datos provenientes desde distintas fuentes”. Finalmente, destaca que “las dos partes mencionadas anteriormente se combinan para dar soporte a todos los requerimientos históricos, analíticos y de inteligencia de negocios”. (37)

### 5.2.15 Visual Studio

Visual Studio es “un conjunto de herramientas y tecnologías asociadas al desarrollo software, que tiene por principal función crear aplicaciones cumpliendo altos estándares de rendimiento, posibilitando a los usuarios generar aplicaciones web, así como otros servicios en entornos diferentes”. (38)

Por otro lado, Danysoft destaca que Visual Studio es “un entorno de desarrollo integrado, la que posee una gran cantidad de atributos que dan apoyo a los desarrolladores de software”, donde de igual manera, destaca que “el IDE de Visual Studio permite editar, depurar y compilar código y, después, publicar una aplicación.”. (39)

### 5.2.16 Matriz RACI

La matriz RACI según DAMA hace referencia a las designaciones de “responsable, explicable, consultado, e informado (RACI)”, donde permite “ayudar a aclarar las funciones, la designación de funciones, responsabilidades, y roles dentro de las distintas áreas del organismo y/o empresa”. (40)

De igual manera, IBM la define como una “manera de garantizar entendimiento de las respectivas asignaciones de roles y responsabilidades dentro de la organización” donde se enmarca como “una manera común para proponer un marco de decisión para aclarar los roles y responsabilidades dentro de los pasos que integran un proceso.”. (41)

### 5.2.17 Trigger

Según Microsoft, un trigger es “un tipo especial de procedimiento almacenado que se gatilla automáticamente cuando se genera un evento en el servidor de bases de datos, es decir, cuando un usuario realiza alguna modificación de registros mediante un evento de lenguaje de manipulación de datos, y en particular, acciones como INSERT, DELETE o UPDATE”. (42)

### 5.2.18 JavaScript

Según Marco Villamar, Javascript es “un lenguaje que sustenta más de una estructura de programación y requiere ser interpretado para eventualmente ser ejecutado”, donde por otro lado, también destaca que los “códigos JavaScript permiten agregar funciones novedosas para ser finalmente ejecutados por el usuario de forma normal, posteriormente a la acción que se va realizar, siendo estas consideradas como un evento específico.”. (43)

Por otro lado, según Rafael Menéndez, JavaScript es “un lenguaje de programación a base de scripts (entendido por ser una consecución de comandos), la cual tiene la particularidad de estar orientado a objetos”, donde también destaca que “los scripts descritos en JavaScript son

mayormente interpretados”, y de esta manera, al ser “JavaScript un lenguaje interpretado, se hace necesario contar con un intérprete para ejecutar código JavaScript, tal como un navegador web”. (44)

### 5.2.19 Node.JS

Según Gabriel Rodríguez, Node.JS “es concebido como un entorno de ejecución para el lenguaje JavaScript, el que esta principalmente orientado a eventos, diseñado para desarrollar aplicaciones en red escalables y de rápida comunicación”. (45)

Por otro lado, según Ignacio Aldarabi, Node.js es un “entorno que se encarga de ejecutar código JavaScript dentro del servidor” donde “gestiona las respectivas tareas de manera asíncrona, lo cual es un concepto revolucionario e innovador, ya que gestiona las ejecuciones manera mucho más eficiente debido a que todo el script se ejecuta en paralelo”. (46)

### 5.2.20 Formato JSON

Según David Barrera, las siglas JSON “significan “notación de objetos JavaScript” y hace referencia a un formato de texto que posibilita la transferencia de datos, siendo este formato fácil de interpretar para los softwares en general, y también de generar archivos con dicho formato”. (47)

## 5.3 *Frameworks* teóricos de apoyo

Para la modelación y posterior implementación de un plan de gobernanza de datos, se hace necesario poder orientar las decisiones en función de experiencias de éxito, y siendo más específico, basarse en *frameworks* que direccionen buenas prácticas de gestión dentro de la institución. Es así como actualmente existe una gran cantidad de *frameworks* disponibles para la implementación del gobierno de datos, los que difieren entre ellos en los respectivos enfoques y pasos que cada una de ellas aborda. La utilidad de los *frameworks* dentro de la presente memoria estará ligada a poder sustentar teóricamente las definiciones presentes en la memoria, además de poder guiar de una manera más estructurada la implementación.

Es así que, para poder abordar el modelo y posterior implementación del gobierno de datos en la Universidad de Chile, se barajaron 4 *frameworks* diferentes en los que apoyar la metodología, en particular los *frameworks* ofrecidos por IBM, DAMA, DGI, y ORACLE, los que exponen distintos enfoques y soluciones de gestión. A continuación, se expondrá el *framework* elegido para la presente memoria, y sus respectivas razones.

### 5.3.1 DAMA

DAMA es una asociación internacional para profesionales de la gestión de datos, donde su principal objetivo consiste en ofrecer lineamientos para el desarrollo de un nuevo paradigma de gestión de los datos, transformándose en referencia para una gran cantidad de organizaciones y profesionales que se desenvuelven en la gestión de datos, aportando recursos, lineamientos, estructuras, y conocimiento sobre la materia. El *framework* DAMA converge en su libro DAMA-DMBOK,

donde definen una visión estándar de industria e instituciones para la implementación de las respectivas funciones ofrecidas para gestión de datos, la terminología teórica y las mejores acciones, sin detallar métodos y técnicas estructuradas y específicas a implementar. (DAMA Internacional, 2009)

DAMA enmarca la gestión de datos dentro de un gobierno de datos robusto, que puedan abarcar gran cantidad de enfoques que se plantean en el *framework*. Es así como DAMA plantea la función de gestión global de datos, abarcando 10 principales funciones: (DAMA Internacional, 2009)

- Gobierno de Datos: Planificación, supervisión y control de la gestión y el uso de datos. (48)
- Gestión de Arquitectura de Datos: Definición del modelo para abordar la arquitectura del dato. (48)
- Desarrollo de datos: Levantamiento, análisis, modelamiento, implementación, prueba, despliegue de estrategias tecnológicas y mantenimiento en función del tiempo. (48)
- Gestión de Operaciones de Datos: Proporcionar apoyo desde la obtención de datos hasta su depuración. (48)
- Gestión de la Seguridad de datos: Garantizar la seguridad, confidencialidad y el acceso adecuado y oportunos de datos. (48)
- Gestión de la Calidad de Datos: Definición, monitorización y mejora de la calidad de datos dentro de los datos a gobernar. (48)
- Gestión de inteligencia de negocios y almacenamiento de datos: Habilidad de informes y análisis que permitan visualizar de manera eficiente los datos. (48)
- Gestión de metadatos: Integrar y proporcionar metadatos dentro de la data existente. (48)
- Gestión de Datos maestros y de referencia: gestión de versiones originales y réplicas. (48)
- Gestión de contenidos y Documentación: gestión de los datos que se encuentran fuera de las bases de datos. (48)

Tal como detalla DAMA, el “Gobierno de Datos está en el corazón de la gestión de los activos de datos. En la representación de las funciones principales de la gestión de datos, dentro de las diez introducidas, el Gobierno de Datos se muestra al medio de estos” donde además destaca que “otra forma de representar la posición de control del Gobierno de Datos es como “el techo de la gestión” sobre otras funciones que ofrece la gestión de datos” (49) como se muestra en la Figura 9.



Figura 9 Casa del data management (DAMA, 2009)

Es así como, en específico, DAMA plantea las siguientes actividades para una correcta implementación del gobierno datos en cualquier organización:

- Comprender las necesidades estratégicas de datos empresariales: Se analizan las actuales necesidades estratégicas de datos de la institución o empresa, además de entender los respectivos alcances de estas. (DAMA Internacional, 2009)
- Desarrollar y mantener la estrategia de datos: Es indispensable poder desarrollar una estrategia transversal que pueda alinearse con las necesidades de la respectiva organización, y que sea sustentable tanto en el corto como largo plazo. (DAMA Internacional, 2009)
- Establecer las funciones de gestión de datos de profesionales y organizaciones: Para esta actividad, es fundamental poder determinar las funciones y responsabilidades de cada área dentro del gobierno de datos, además de designar al personal técnico que se encargará de la gestión de datos, y las acciones técnicas que se tomen dentro del desarrollo del plan. (DAMA Internacional, 2009)
- Establecer organizaciones, administración del gobierno de datos, y roles de profesionales de datos: Se debe establecer quiénes serán los integrantes del Consejo Gobierno de datos (Estratégico), Comité de administración de datos (Comité táctico de gobierno de datos), y Equipos de administración de datos (Mesas de trabajo con áreas involucradas). (DAMA Internacional, 2009)
- Revisión y aprobación de arquitectura de datos: Se levanta toda la arquitectura actual que mantiene la organización con respecto a los datos, y se aprueban los lineamientos para las eventuales acciones que se tomarán en función de sostener la implementación de gobernanza de datos. (DAMA Internacional, 2009)
- Estimar los valores de activos de datos: Se genera las estimaciones del valor de los activos relacionados al dato a gobernar, además de levantarlos e identificarlos oportunamente (DAMA Internacional, 2009)

- Administrar y resolver problemas relacionados con datos: Luego de la previa identificación y levantamiento de la respectiva información relacionada al dato a gobernar, se procede a resolver los problemas levantados. (DAMA Internacional, 2009)

Es así como el *framework* DAMA se presenta con una gran robustez en las aristas abordadas, que, si bien posee una gran cantidad de enfoques para la implementación de esta, permite una flexibilidad mayor que los demás *frameworks*. Por otro lado, cabe destacar que DAMA presenta un capítulo completo enfocado a la recomendación de acciones y buenas prácticas para la gestión de calidad de los datos, arquitectura del dato, y la Metadata, enfoques que la Universidad de Chile debe abarcar para solucionar los problemas que actualmente se presentan a nivel institucional. Finalmente, el *framework* ofrecido por DAMA ha servido como referencia para una gran cantidad de proyectos de gobierno de datos, existiendo mucha información montada en la web, además una comunidad de ayuda para apoyar la implementación, permitiendo de esta manera brindar un soporte importante para la búsqueda de información crucial para su respectiva implementación.

### 5.3.2 Guía técnica MINTIC

Tal como se mencionaba en la sección anterior, la metodología DAMA ha servido como guía para muchos proyectos de gobernanza internacionalmente, existiendo marcos de trabajo asociadas a la metodología, los que permiten guiar de mejor manera la proyección que ofrece DAMA. Es así como el actual gobierno de datos diseñado para la Universidad de Chile estará basado en una de las experiencias internacionales de gobierno más importantes dentro de Latinoamérica; la guía técnica que ofrece el MinTic. El MinTic es el ministerio de tecnologías de la información y la comunicación de Colombia, la que “se encarga de diseñar, adoptar y promover las políticas, lineamientos, planes, programas y proyectos del sector de las Tecnologías de la Información y las Comunicaciones.” (50)

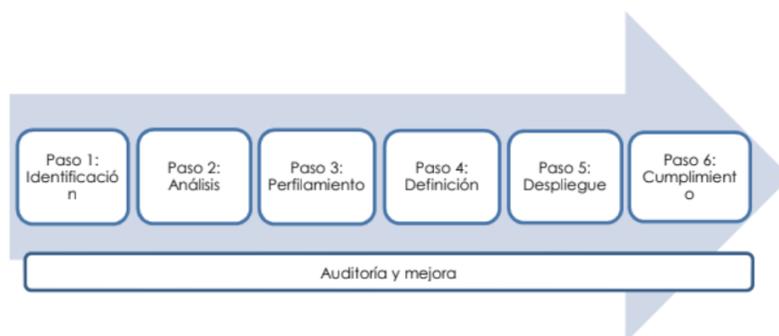
MinTic ofrece una guía técnica que “apoya la implementación de lineamientos asociados al registro y seguimiento de información, al establecimiento de los mecanismos de monitorización, la creación y mantenimiento del repositorio transversal de los activos de información involucrados sobre los datos a gobernar, y consolidación de los datos dentro de la organización, para presentar las acciones sobre la gestión que pueden ser aplicados por las organizaciones para la implementación del gobierno del dato”. (51)

El marco de trabajo propuesto define su enfoque en los siguientes términos explicitados en la figura 10.



Figura 10 Funciones de la gestión de datos (María Isabel Mejía Jaramillo, 2014)

Donde en cada uno de estos ámbitos se busca garantizar para el dato una definición a nivel institucional, y así organizar el uso eficiente del mismo dentro de la organización. Estos diferentes ámbitos deben mantener una visión transversal con la misión de la universidad y el área TI, los que deben contar con roles y procesos específicos de Gobierno del Dato. (María Isabel Mejía Jaramillo, 2014). MinTic entabla la implementación del gobierno de datos en seis pasos fundamentales. En la figura 11 se muestra una referencia gráfica de estos pasos.



*Figura 11 Pasos implementación (María Isabel Mejía Jaramillo, 2014)*

A partir de la figura 11 se ve cómo la estrategia de gobierno del dato inicia con la identificación de los datos a gobernar y la definición de los alcances, avanza con el análisis del estado del dato dentro de flujos y procesos dentro de la institución, continúa con la definición detallada de elementos que se considerarán para gobernar, progresa con el establecimiento y comunicación de estos elementos en función de soluciones tecnológicas, y finaliza con la verificación de cumplimiento alrededor del gobierno del dato. En forma transversal y continua, la estrategia planteada debe estar constantemente realizando auditorías y mejoras para el dato en cuestión. Esta estrategia de gobierno utiliza datos maestros, metadatos funcionales, responsabilidades en gobierno y data, mecanismos de acceso a la información, datos orquestados entre sistemas y normativas para codificación, estandarización y unificación, a través del ciclo de vida del dato. (María Isabel Mejía Jaramillo, 2014)

A partir de los pasos enunciados por la guía de gobierno de datos que ofrece el MINTIC para la implementación de este, para efectos de esta memoria se tomaron 4 de estos pasos; Identificación, análisis, perfilamiento y despliegue. En cuanto al paso de definición, en términos de alcance para esta memoria, será excluido dentro de los pasos pactados para la implementación de la gobernanza de datos en la Universidad de Chile, debido a que existe una implicancia de las demás áreas que toman decisiones sobre datos dentro de la universidad, existiendo muchas veces problemas a la hora de definir y alcanzar acuerdos con estas, definiendo normativas de codificación, nuevos accesos a bases de datos transaccionales, reglamentación, seguridad, entre otras.

Finalmente, para el paso de cumplimiento, tampoco entra dentro de los alcances propuestos para esta memoria, debido a la naturaleza inherente que tiene el gobierno de datos, la que hace referencia a poder verificar el cumplimiento periódico de todas las estrategias al largo plazo, después de la respectiva implementación, lo que entra dentro de lo definido para el trabajo futuro.

## 6 ALCANCES

Sin duda, cuando se habla de gobierno de datos hacemos referencia a un cambio de paradigma en cuanto a cómo vemos los datos y cómo estos fluyen en flujos y procesos de la Universidad de Chile. Es así como la implementación de un proyecto de esta magnitud debe ser cauto en cuanto al alcance que este tendrá en sus iteraciones iniciales, para así ir progresando poco a poco a más aristas dentro de la institución, todo contemplando un proyecto estratégico que proyecta decisiones tanto a corto como a largo plazo.

Para efectos que tendrá este plan de gobernanza en la Universidad de Chile pactado para esta memoria, y dentro de lo expuesto por el *framework* DAMA y el plan de trabajo presentado por el MINTIC, se gobernarán datos con un especial enfoque en gestión de calidad del dato, además de abarcar arquitectura y metadata de los datos a gobernar, entendiendo las necesidades más importantes dentro de la gestión y administración de la universidad.

Por otro lado, actualmente en las bases de datos transaccionales *SUG* y *WEB* (Bases más pobladas y utilizadas por distintas aplicaciones) existen una gran cantidad de tablas y/o entidades, por lo que para efectos de esta memoria se considerará los datos albergados dentro de las tablas “INDIVIDUO”, “DOTACION” y “DATOS\_PERSONALES”, estando las dos primeras en *SUG* y la última en *WEB*, siendo estas las tablas más pobladas y relevantes dentro de la universidad, debido a que están conectadas a gran parte de las aplicaciones de ingreso de datos que hoy en día la Universidad de Chile posee en funcionamiento.

Finalmente, este gobierno de datos posee características iterativas, y en función de esto, el alcance de esta memoria estará acotado a poder gobernar los atributos “Nombre”, “Apellido Materno” y “Apellido paterno”, levantando así documentación que hoy en día no se posee a nivel institucional, democratizando el conocimiento de los actuales procesos y flujos de la universidad. Por otro lado, se dispondrá indicadores y estándares, además de herramientas tecnológicas para poder monitorizar estos atributos dentro de las tablas más pobladas de la universidad de Chile, y así poder abordar tanto problemas superficiales como de raíz.

## 7 METODOLOGIA

La metodología propuesta para la implementación del gobierno de datos en la Universidad de Chile comprende un sustento teórico basado en el *framework* DAMA, donde se permite obtener acciones, lineamientos y prácticas para utilizar dentro de la misma. Por otro lado, la estructura de la metodología en términos de pasos que comprende esta memoria está basada en la guía que ofrece el MINTIC sobre la implementación de gobernanza de datos, la que es adaptada en función de los alcances y las necesidades más importantes que posee hoy en día la institución en términos de gestión y administración de datos.

El plan propuesto para la implementación constará de 2 etapas, demarcando en primer lugar (Etapa 1) los parámetros organizativos propuestos para la disposición de las respectivas autoridades que regirán como tomadores de decisiones dentro de la ejecución del gobierno de datos, correspondiendo al eje organizativo del plan. La segunda parte (Etapa 2) estará enfocada en el detalle de los respectivos pasos a seguir para la implementación técnica de este plan de gobierno

sobre los datos de la Universidad de Chile, donde esta inicia con la identificación y definición de los datos a gobernar, para luego generar un estudio del estado en el que actualmente se encuentran estos datos sobre aplicaciones, continúa con la definición de estándares e indicadores de calidad asociados al dato a gobernar, y finaliza con la implementación de las soluciones tecnológicas pactadas para abordar los problemas. En forma transversal, los pasos a seguir luego de la implementación contemplan una continua realización de auditorías y posteriores mejoras para el dato gobernado. En la figura 12 se despliega el respectivo diagrama de etapas y fases que comprenden la metodología de este proyecto.



*Figura 12 Etapas metodología*

## 7.1 Etapa 1: Conformación comités y mesas de trabajo

La etapa 1 de la implementación del plan de gobernanza de datos en la Universidad de Chile corresponde al eje organizativo del mismo, radicando la importancia de este en función de poder definir y generar dos comités, uno con atribuciones aprobatorias de acciones sobre la gobernanza, es decir, que permita aprobar las acciones sobre los datos a gobernar, y otro con atribuciones propositivas, que pueda proponer acciones sobre los datos a gobernar. Por otro lado, también se define la estructura de mesas de trabajo, las que tienen por función levantar la información relevante para la institución, además de brindar apoyo en la toma de decisiones sobre los datos. Esta etapa 1 de la gobernanza está ligada directamente al cumplimiento del primer objetivo específico pactado para la presente memoria, y en particular, poder establecer los respectivos lineamientos organizativos para una correcta implementación del plan de gobernanza, generando las bases tomadoras de decisiones para la acción tomada dentro de la presente memoria.

La conformación de un comité estratégico de gobierno de datos es la primera actividad a realizar durante la etapa 1. En dicho comité se congrega a las máximas autoridades de la institución, incluyendo la participación de académicos, estudiantes y funcionarios, teniendo la función y responsabilidad de aprobar o rechazar las decisiones tomadas por el comité táctico. Por otro lado, se define la estructura del comité táctico, el que tiene la labor de definir y levantar todas las acciones que se tomarán en cada uno de los pasos proyectados para este plan, realizando un continuo trabajo con las distintas mesas de trabajo formadas por las respectivas áreas del negocio involucradas en los datos a gobernar. Finalmente se define la estructura de las mesas de trabajo, congregando a todas las áreas técnicas que tengan implicancia en el dato a gobernar, para así levantar información fundamental sobre los respectivos datos, además de apoyar la toma de decisiones sobre medidas que se apliquen sobre ellos.

## 7.2 Etapa 2: Implementación técnica gobernanza

Dentro de esta etapa se detallan todos los pasos técnicos a seguir para completar una correcta implementación del gobierno de datos en la Universidad de Chile. Los pasos tendrán relación con lo expuesto por el MINTIC, siguiendo una estructura lógica similar, entendiendo las necesidades que comprometen a la Universidad de Chile, y moldeando la implementación en función de estas. Cabe destacar que, para cada paso, se considerarán entradas y salidas, donde las primeras harán referencia a “inputs” necesarios para la realización de estos, y las segundas a los entregables finales que se dispondrán luego de completar cada uno de estos pasos.

### 7.2.1 Paso 1: Identificación de datos maestros y/o importantes, definición de metadatos funcionales

Esta etapa tiene la labor identificar los datos a gobernar, y la definición de metadatos funcionales con respecto a los datos elegidos para las primeras iteraciones del gobierno de datos. Es importante destacar que se busca trabajar sobre un subconjunto de información inicial que debiese considerar los datos importantes para la universidad, esto con el fin de poder afianzar el concepto de gobierno y realizar labores que entreguen resultados medibles. El abarcar todos los datos en una primera fase es de alto riesgo, ya que los respectivos análisis y mediciones de desempeño serían a mayor escala, dificultando el correcto manejo de éstos. Las actividades que se llevan a cabo para cumplir con los requerimientos de esta etapa son las siguientes:

- Identificación datos a gobernar.
- Realización de un inventario de activos, levantamiento de activos tecnológicos como de datos asociados el dato a gobernar.
- Identificación metadatos para los datos a gobernar.
- Identificación de responsables sobre pasos del plan.

Posterior a la definición de actividades, se plantearon respectivas entradas y salidas para materializar la realización de cada paso. Estas se detallan en la figura 13.

Paso 1: Identificación de datos maestros y/o importantes, definición de metadatos funcionales	
<b>OBJETIVO GENERAL</b>	Identificar los datos maestros y definir los metadatos de cada uno de ellos.
<b>ENTRADA</b>	<b>SALIDA</b>
<ul style="list-style-type: none"> <li>• Inventario de activos de información de la institución.</li> <li>• Colaboración áreas técnicas involucradas.</li> </ul>	<ul style="list-style-type: none"> <li>• Identificación datos a gobernar.</li> <li>• Caracterización de metadatos para el dato a gobernar.</li> <li>• Creación de matriz responsabilidades.</li> </ul>

Figura 13 Entrada y salidas paso 1

## 7.2.2 Paso 2: Revisión de tratamiento de datos a gobernar dentro de flujos de la universidad

Para esta etapa se debe realizar una correcta revisión sobre tratamiento de datos a gobernar. Se debe hacer un foco en cuanto a la revisión de la granularidad de la información con respecto al acceso a las bases de datos, y la definición de flujos de datos por sobre aplicaciones que ingresan datos a las bases de datos transaccionales de la universidad. Las actividades que se llevarán a cabo para cumplir con los requerimientos de esta etapa serán los siguientes:

- Levantamiento en procesos y flujos dentro de aplicaciones, para así consolidar una visión única del dato a gobernar.
- Creación de matriz CRUD.

Posterior a la definición de actividades, se plantearon respectivas entradas y salidas para materializar la realización de cada paso. Estas se detallan en la figura 14.

Paso 2: Análisis, Revisión de tratamiento de datos maestros y/o importantes	
<b>OBJETIVO GENERAL</b>	<ul style="list-style-type: none"> <li>• Revisión del tratamiento de datos maestros en procesos y flujos dentro de aplicaciones que recogen datos.</li> </ul>
<b>ENTRADA</b>	<b>SALIDA</b>
<ul style="list-style-type: none"> <li>• Caracterización de los datos a ser gobernados</li> <li>• Apoyo de áreas funcionales</li> <li>• Apoyo de áreas de desarrollo y mantención de VTI</li> </ul>	<ul style="list-style-type: none"> <li>• Matriz CRUD</li> <li>• Mapeo de flujos de datos dentro de aplicaciones que posee la Universidad de Chile</li> </ul>

*Figura 14 Entrada y salidas paso 2*

## 7.2.3 Paso 3: Perfilamiento

Este perfilamiento estará ligado al levantamiento de los respectivos indicadores de calidad que se proponen en el modelamiento del plan, además de la definición de estándares de datos que se cumplirán a modo de visión única dentro de la universidad. Las actividades que se llevarán a cabo para cumplir con los requerimientos de esta etapa serán los siguientes:

- Definición de estándares de cumplimiento para datos a gobernar, explicitando el formato correcto a congeniar para una visión única del dato.
- Elección y definición de indicadores de calidad a utilizar para poder medir la respectiva calidad de los datos en las bases transaccionales de la universidad.

Posterior a la definición de actividades, se plantearon respectivas entradas y salidas para materializar la realización de cada paso. Estas se detallan en la figura 15.

Paso 3: Perfilamiento	
<b>OBJETIVO GENERAL</b>	Perfilamiento de datos, de información y de la función institucional.
<b>ENTRADA</b>	<b>SALIDA</b>
<ul style="list-style-type: none"> <li>• Identificación de flujos de datos entre sistemas.</li> <li>• Administradores de bases de datos y equipos de desarrollo VTI y otras unidades</li> </ul>	<ul style="list-style-type: none"> <li>• Estandarización de datos a gobernar, generando respectivos formatos que se usarán a nivel institucional</li> <li>• Definición de indicadores de calidad a considerar dentro de las métricas de monitorización</li> </ul>

Figura 15 Entrada y salidas paso 3

### 7.2.4 Paso 4: Despliegue

Luego de poder haber levantado la arquitectura actualizada, responsables, estándares de datos y los respectivos indicadores a utilizar, se hace necesario poder desplegar las soluciones que permitan tanto monitorizar como generar limpieza dentro de las bases institucionales de la universidad. En esta etapa se podrá desplegar las herramientas tecnológicas e implementar soluciones de monitorización y limpieza para el gobierno de datos. Las actividades que se llevarán a cabo para cumplir con los requerimientos de esta etapa serán los siguientes:

- Construcción de *dashboard* dinámico de calidad, cumpliendo con los respectivos estándares e indicadores de calidad definidos en el paso anterior.
- Analizar resultados de indicadores de calidad sobre los atributos y tablas a gobernar.
- Ejecución de acciones correctivas y de limpieza para el dato a gobernar.

Posterior a la definición de actividades, se plantearon respectivas entradas y salidas para materializar la realización de cada paso. Estas se detallan en la figura 16.

Paso 4: Despliegue	
<b>OBJETIVO GENERAL</b>	Despliegue de la arquitectura de VTI e implantación de soluciones para gobierno de datos.
<b>ENTRADA</b>	<b>SALIDA</b>
<ul style="list-style-type: none"> <li>• Arquitectura TI.</li> <li>• Inteligencia de negocios (BI).</li> </ul>	<ul style="list-style-type: none"> <li>• Dashboard de calidad (Monitorización)</li> <li>• Informe de resultados por tablas y atributos</li> <li>• Limpieza de datos en función de indicadores críticos.</li> </ul>

## 8 IMPLEMENTACION

### 8.1 Etapa 1: Definición estructura y funciones comités y mesas de trabajo

Tal como se detalla en la metodología propuesta, la primera etapa de implementación de gobernanza de datos parte con la estructuración de los respectivos comités, definiendo tanto objetivos como funciones que tendrán. Para la presente memoria, es de importancia poder definir estos comités, dado que gran parte de las acciones tomadas para la respectiva implementación son levantadas en el comité táctico, para posteriormente ser aprobadas en el comité estratégico, y finalmente dar cumplimiento a estas.

#### 8.1.1 Comité estratégico de Gobierno de Datos

El Comité Estratégico de Gobierno de Datos estará conformado por el prorector, el que asumirá el rol de la dirección de este comité, en conjunto con las vicerrectorías, donde el vicerrector de la VTI asumirá como subdirector del comité. Por otro lado, también se incluirá un representante del senado universitario, un delegado/a del estamento académico, un delegado/a de los estudiantes, y un delegado/a del personal de colaboración, quienes serán designados por las organizaciones representativas de sus respectivos estamentos. La respectiva resolución de creación del comité estratégico de gobierno de datos se encuentra en el anexo A.

De esta manera, los objetivos del comité estratégico de gobierno de datos son los siguientes:

- Supervisar el desarrollo del programa de gobierno de datos a nivel estratégico.
- Aprobar definiciones presentadas sobre reglas, flujos, políticas de datos, presentados por el nivel táctico.
- Resolver problemas que no pueden solucionarse por diferencias de visión en las mesas de trabajo de los equipos operativos.
- Comunicar y promover el programa a los diferentes niveles de la organización.

Por otro lado, las funciones que tendrá este comité estratégico de gobierno de datos sobre la acción del plan de gobernanza serán las siguientes:

- Gestión para el desarrollo y avance del plan de gobierno de datos. Definiendo orientaciones específicas que sean necesarias a lo largo del desarrollo del plan de trabajo.

- Proponer y fijar plazos para la implementación de las decisiones relacionadas al plan de gobierno de datos.
- Verificación de los resultados del proceso.
- Mediar o decidir ante problemáticas existentes en las mesas de trabajo que no han podido ser resueltas sobre responsabilidades o decisiones sobre los datos.

### 8.1.2 Comité táctico de Gobierno de Datos

El objetivo del comité táctico de Gobierno de datos es definir las reglas, y procesos del programa, además de poder garantizar que estos requerimientos se cumplan con respecto a la aprobación del comité estratégico. Por otro lado, también deben evaluar los requisitos y problemas con los respectivos datos a gobernar. Por último, posee el objetivo de medir e informar acerca de la calidad de los datos y documentación de estos. Este comité táctico se compondrá únicamente por la dirección de datos de la Universidad de Chile. En el anexo B se expone la estructura de los comités. De esta manera, los objetivos del comité táctico de gobierno de datos son específicamente los siguientes:

- Levantar documentación implicada en todos los datos a gobernar.
- Identificar y documentar las características de cada uno de los datos que se manejan en los procesos, además de participar activamente en perfilar, revisar y aprobar definiciones y formatos de los respectivos elementos del dato estandarizado.
- Crear, actualizar, eliminar y mantener datos basados en los estándares definidos, además de establecer qué usuarios pueden crearlos, visualizarlos, editarlos o eliminarlos.
- Describir las políticas, fórmulas, reglas de negocio y la forma en que se le dará mantenimiento a la información.
- Identificar e implementar oportunidades de mejora en la respectiva calidad de los datos a gobernar.

### 8.1.3 Mesas de trabajo de Gobierno de Datos

El objetivo principal de la mesa de trabajo es la de levantar la información relevante para la institución, además de brindar apoyo en la toma de decisiones de estándares de datos, logrando así acordar el mismo uso para todos dentro de la institución. También posee la función de identificar y documentar las características de los datos que se manejan en los respectivos procesos.

Finalmente, esta mesa tiene por objetivo determinar accesos sobre la información, además de describir reglas de negocio y forma de mantención de los datos.

La conformación de las mesas de trabajo variará de acuerdo con los datos que se estén trabajando, pero debe contar con representantes de unidades dueñas de los datos a trabajar, representantes de la oficina de seguridad, representantes de la dirección de datos, y representantes de desarrollo TI, esto para conformar una mesa inicial de labores. De acuerdo con la necesidad de avanzar con el gobierno del dato se puede hacer necesario agregar nuevos integrantes o realizar modificaciones en la conformación del grupo. Estas mesas de trabajo seguirán la estructura mostrada en el anexo C.

Finalmente, luego de la conformación de los respectivos comités y las mesas de trabajo, se da cumplimiento al primer objetivo específico pactado para esta memoria, estableciendo los lineamientos organizativos para la correcta implementación de la gobernanza de datos dentro de la Universidad de Chile.

## 8.2 Etapa 2: (Paso 1) Identificación de datos maestros y/o importantes, definición de metadatos funcionales

Para la realización de este paso se definieron entradas y salidas fundamentales para poder generar un mejor entendimiento en cuanto a la realización del paso. La definición de las respectivas entradas y salidas de este paso se definen en la figura 13.

### 8.2.1 Entrada: Identificación datos a gobernar y mesas de trabajo

Una de las claves en una correcta definición de alcances y proyecciones dentro de cualquier plan de gobierno de datos, es la iteración progresiva de estos, de tal forma que se pueda abarcar de buena manera estas definiciones proyectadas. Es así que, como comité táctico de gobierno de datos, con una posterior aprobación del comité estratégico, se aprueba que en las primeras iteraciones del plan de gobierno de datos, los datos a gobernar serán los que corresponden a datos personales de funcionarios, académicos, y alumnos de la institución, y en específico, poder abarcar completamente el trabajo de gobierno en la primera iteración sobre los atributos "Nombre", "Apellido Materno", y "Apellido Paterno", dentro de las tablas "DOTACION", "DATOS PERSONALES", e "INDIVIDUO". Esta priorización de datos a gobernar se generó en función de la importancia que tiene este dato dentro de procesos cruciales para la gestión y administración, tales como matrícula, documentos de título, emisiones de contratos, entre otros, además de la transversalidad que tiene este dato dentro de la universidad, tanto en estos procesos como en aplicaciones. Posterior a la definición de estos atributos a gobernar, la decisión final se lleva a cabo por el comité estratégico de datos, dando aprobación y finalmente la iniciación oficial del proceso de gobernanza dentro de la Universidad de Chile.

Luego de la priorización y definición de atributos a gobernar, se proyecta la conformación de las mesas de trabajo, estructurando el comité táctico de gobierno de datos. Posteriormente, el proyecto se desenvuelve en un continuo trabajo dentro de la dirección de datos, en conjunto con áreas involucradas dentro de la gestión y administración de la Universidad de Chile, formando mesas de trabajo en función de levantar información sobre los datos a gobernar, considerando niveles de problemas operativos, y la importancia dentro de la función de sus respectivas actividades. El

formato y un ejemplo de conformación de las respectivas mesas de trabajo se muestra en el anexo D.

Dentro de las respectivas mesas de trabajo se integraron distintas áreas involucradas en los datos a gobernar, considerando consumidores del dato, custodios de los datos, generadores del dato, y desarrolladores. Si bien los alcances de esta memoria consideran gobernar el atributo "Nombre", "Apellido Materno", y "Apellido Paterno", para esta categorización se considera todos los datos personales destinados a gobernar dentro de las primeras iteraciones del gobierno de datos en la Universidad de Chile, luego de la respectiva implementación pactada para esta memoria.

## 8.2.2 Entrada: Inventario de Activos

Por otro lado, una de las entradas determinantes para poder generar una visión holística de la infraestructura de activos que tiene la universidad, consiste en la elevación de un completo inventario de activos institucionales correspondiente al dato a gobernar. Para esto se estructura una plantilla Excel en donde se consideran las siguientes clasificaciones para caracterizar los respectivos activos en cuestión; nombre del activo, descripción del activo, contenedor en base de datos si corresponde, tipo de activo, propietarios, responsable del activo, custodio, y la respectiva utilidad del activo en función de la información que provee. El correspondiente inventario se encuentra expresado en las tablas 2 y 3.

Nombre activo	Descripción del activo	Contenedor en base de datos	Tipo de activo	Propietario
Individuo	Entidad en base de datos	SUG	Datos	VTI
Persona	Entidad en base de datos	Personas	Datos	VTI
Dotación	Entidad en base de datos	SUG	Datos	VTI
Datos peronales	Entidad en base de datos	SRVPASP	Datos	VTI
cu_alumnos	Entidad en base de datos	SUG	Datos	VTI
Diplomas	Entidad en base de datos	Diploma electronico	Datos	VTI
PCO_observacioncertificado	Entidad en base de datos	Certificados	Datos	VTI
PCO_solicitante	Entidad en base de datos	Certificados	Datos	VTI
aut_username	Entidad en base de datos	Authorization	Datos	VTI
Consulta	Entidad en base de datos		Datos	VTI
SUG	Base de datos		Datos	VTI
WEB	Base de datos		Datos	VTI
Orauchdbprod	Base de datos		Datos	VTI
SQLprod	Base de datos		Datos	VTI
Mongoprimary6401	Base de datos		Datos	VTI
FOCES.INCIDENTE	Entidad en base de datos		Datos	VTI
FOCES.INTEGRANTE	Entidad en base de datos		Datos	VTI
FOCES.INTEGRANTEFS	Entidad en base de datos		Datos	VTI
OSTGRADO.CARTA_REC_FICHA	Entidad en base de datos		Datos	VTI

Tabla 2 Inventario de activos

Responsable del activo	Custodio	Utilidad de la información
VTI	Dirección de datos	Información personal de alumnos, funcionarios, academicos, y empresas
VTI	Dirección de datos	Información personal de alumnos, funcionarios y academicos
VTI	Dirección de datos	Información sobre vinculo contractual de personas con la Universidad de Chile
VTI	Dirección de datos	Información de las cuentas pasaporte
VTI	Dirección de datos	Información de alumnos pertenecientes a la Universidad de Chile
VTI	Dirección de datos	Información

Tabla 3 Inventario de activos (Tabla complementaria a la previa)

Por otro lado, para cada activo clasificado y levantado, se valoriza con respecto a 3 definiciones; confidencialidad, integridad y disponibilidad. Para la respectiva valorización se consideraron 4 niveles de selección para cada definición, tal como se detalla en la figura 17 ,

Confidencialidad	Alta (3)	Información disponible sólo en procesos específico, que no debe ser conocida por ninguna persona sin autorización previa, lo que implicaría en un impacto negativo dentro de la organización
	Media (2)	Información disponible en gran parte de los procesos, que de ser conocida por una persona sin autorización previa, implicaría en un impacto negativo dentro de la organización
	Baja (1)	Información que puede ser entregada o publicada sin restricciones de autorización para cualquier persona dentro de la universidad
	NO CLASIFICADA (0)	Activos de Información no clasificados
Integridad	Alta (3)	Información cuya pérdida de exactitud conlleve directamente a repercusiones negativas dentro de la organización
	Media (2)	Información cuya pérdida de exactitud puede conllevar a repercusiones negativas dentro de la organización
	Baja (1)	Información cuya pérdida de exactitud no conlleve directamente a repercusiones negativas dentro de la organización
	NO CLASIFICADA (0)	Activos de Información no clasificados
Disponibilidad	Alta (3)	La no disponibilidad del activo conlleva directamente a repercusiones negativas dentro de la organización
	Media (2)	La no disponibilidad del activo puede conllevar a repercusiones negativas dentro de la organización
	Baja (1)	La no disponibilidad del activo no conlleva a repercusiones negativas dentro de la organización
	NO CLASIFICADA (0)	Activos de Información no clasificados

*Figura 17 Clasificación de activos*

La clasificación de activos de información tiene como objetivo asegurar que los activos reciban los niveles de categorización pertinentes en función de su importancia, además de considerar otras características particulares que requieran algún tipo de manejo particular con respecto a la valorización.

Finalmente, se satisface uno de los puntos más significativos del sistema de gestión de la información, el que permite poder levantar todos los hardware, software, servicios, recursos humanos, entre otros, que eventualmente generan valor como activo en función de los datos a gobernar. Este inventario de activos se trabaja de la mano con áreas implicadas en la custodia, consumo, y responsabilidad, levantando las respectivas aplicaciones que procesan datos, las áreas involucradas en la gestión de estos datos desde las aplicaciones, las tablas que poseen este atributo en las bases de datos, y en general toda la información que sea relevante de levantar como activo institucional relacionado a los atributos "Nombre", "Apellido Paterno" y "Apellido Materno".

### 8.2.3 Salida: Diccionario de metadatos

Luego de un levantamiento transversal de gran parte de los activos de información asociados al dato a gobernar, además de la definición de las respectivas áreas que conformarán las primeras mesas de trabajo, se procede a generar las respectivas actividades destinadas a obtener las salidas enunciadas en la figura 13.

Tal como se enuncia en los respectivos alcances de la presente memoria, se definen criterios para caracterizar la metadata de los atributos "Nombre", "Apellido Materno", y "Apellido Paterno". El levantamiento de metadata se realiza en una plantilla Excel, la cual integra columnas necesarias para poder definir la "información sobre" los respectivos datos a gobernar. Este documento integra información sobre la metadata técnica del dato, como el tipo de almacenamiento dentro de la base

de datos, si es *nulleable* (Si el dato puede ser nulo), *data length* (Tamaño del dato), escala del dato, y la descripción del atributo asociado dentro de las tablas a gobernar, estableciendo de igual manera las respectivas responsabilidades que se tendrán en función de los datos, identificando a los custodios, propietarios, y consumidores de los datos a gobernar. En función de la confidencialidad solicitada para la categorización de responsabilidades, están no serán desplegadas en la tabla 5. La información técnica sobre la metadata de los datos a gobernar fue extraída desde una tabla que posee la información técnica sobre gran parte de las tablas que se alojan en la base de datos "SUG" y "WEB", además de obtener información complementaria en reuniones técnicas con las respectivas áreas involucradas, para así finalmente poder llegar a una visión unificada sobre la descripción del respectivo atributo. Para visualizar de mejor manera el trabajo de esta salida, se despliega la plantilla realizada en las tablas 4 y 5.

TABLE_NAME	COLUMN_NAME	DATA_TYPE	DATA_LENGTH	DATA_PRECISION	DATA_SCALE	NULLABLE	COMMENTS
INDIVIDUO	ID	VARCHAR2	15			N	Codigo del individuo, el cual lo identifica en forma unica
INDIVIDUO	PATERNO	VARCHAR2	128			N	Apellido paterno o razon social
INDIVIDUO	MATERNO	VARCHAR2	50			Y	Apellido materno, para las personas
INDIVIDUO	NOMBRES	VARCHAR2	50			Y	Nombres en el caso de las personas
INDIVIDUO	DIRECCION	VARCHAR2	200			Y	Direccion
DATOS PERSONALES	ID	VARCHAR2	15			N	Codigo del individuo, el cual lo identifica en forma unica
DATOS PERSONALES	PATERNO	VARCHAR2	128			N	Apellido paterno o razon social
DATOS PERSONALES	MATERNO	VARCHAR2	50			Y	Apellido materno, para las personas
DATOS PERSONALES	NOMBRES	VARCHAR2	50			Y	Nombres en el caso de las personas
DATOS PERSONALES	DIRECCION	VARCHAR2	200			Y	Direccion
DOTACION	ID	VARCHAR2	15			N	Codigo del individuo, el cual lo identifica en forma unica
DOTACION	PATERNO	VARCHAR2	128			N	Apellido paterno o razon social
DOTACION	MATERNO	VARCHAR2	50			Y	Apellido materno, para las personas
DOTACION	NOMBRES	VARCHAR2	50			Y	Nombres en el caso de las personas

Tabla 4 Inventario de metadata

COMMENTS	Custodio del dato	Propietario del dato	Consumidores del dato
Codigo del individuo, el cual lo identifica en forma unica			
Apellido paterno o razon social			
Apellido materno, para las personas			
Nombres en el caso de las personas			
Direccion			
Codigo del individuo, el cual lo identifica en forma unica			
Apellido paterno o razon social			
Apellido materno, para las personas			
Nombres en el caso de las personas			
Direccion			
Codigo del individuo, el cual lo identifica en forma unica			
Apellido paterno o razon social			
Apellido materno, para las personas			
Nombres en el caso de las personas			
Direccion			

Tabla 5 Inventario de metadata

## 8.2.4 Salida: Matriz de responsabilidades

Por otro lado, siguiendo con las salidas que fueron proyectadas para este paso, finalmente se tiene la matriz de roles, en la cual se congrega las respectivas responsabilidades que tendrán distintos directivos y áreas implicadas en el negocio con respecto al alcance de las entradas y salidas posteriores de cada paso pactado para la implementación del gobierno de dato en la Universidad de Chile. Esta matriz se realiza en Excel, donde en el eje superior se detallan los roles (áreas implicadas en las mesas técnicas), en el eje vertical las distintas actividades de cada paso, y en la intersección se generó una categorización de roles, considerando; Responsable, Aprueba, Consultado e Informado. La matriz RACI generada para el proyecto se detalla en la figura 18 y 19.

Tal como se expone en esta planilla, se consideran 3 actores dentro del proyecto de gobierno de datos; Mesas de trabajo, Comité estratégico, y el Comité táctico. En general, se puede proyectar

una responsabilidad total en el comité táctico, considerándose estos como los gestores y artífices del plan. Por otro lado, tal como se detallaba en la sección anterior, la toma de decisiones a más alto nivel dentro del proyecto está a cargo del comité estratégico, teniendo principalmente un papel aprobatorio en función de las medidas tomadas. Por último, se destaca la participación de las mesas de trabajo levantadas, que tienen por función principal poder ofrecer decisiones en función de un conocimiento técnico, considerando las necesidades de las respectivas áreas del negocio implicadas, teniendo un rol de consulta dentro del desarrollo del plan de gobierno de datos.

ID	Tarea	Estado	ROLES		
			CT	CE	Mesas trabajo
	Identificación de datos maestros y/o importantes, definición de metadatos funcionales				
	Matriz RACI		R	A	C
	Caracterización de datos maestros de la institución y reglas de validación funcional para el dato a gobernar.		R	A	C
	Caracterización de metadatos para el dato a gobernar.		R	A	C
	Plan de Gobierno de los datos		R	A	C
	Reportes de calidad de los datos y sus acciones de mejora.		R	A	C
	Categorización de los datos de acuerdo a frecuencia y volumen de demanda		R	A	C
	Caracterización de matriz CRUD.		R	A	C
	Mapas de proceso del dato		R	A	C
	Directorio de unidades involucradas en el flujo del dato		R	A	C

Figura 18 Matriz RACI

R	Responsable	Asignado para completar la tarea
A	Aprueba	Tiene autoridad para tomar decisiones finales y rendición de cuentas para su finalización. (solo uno por tarea)
C	Consultado	Un asesor, parte interesada o experto en la materia que es consultado antes de una decisión o acción
I	Informado	Debe ser informado después de una decisión o acción

Figura 19 Desglose matriz RACI

De esta manera, luego de identificar los respectivos datos a gobernar para efectos de esta memoria, y de haber levantado el inventario de activos, la matriz de responsabilidades, y el diccionario de metadatos en función de estos, se da cumplimiento al segundo objetivo específico plasmado para esta memoria.

### 8.3 Etapa 2: (Paso 2) Revisión de tratamiento de datos a gobernar dentro de flujos de la universidad

Luego de poder identificar los datos a gobernar dentro del plan, y levantar información sobre los respectivos activos que están involucrados con estos datos a gobernar, además de plasmar las responsabilidades que se tendrán dentro del proyecto, se procede a poder levantar la arquitectura de los datos involucrados en la gobernanza.

Las entradas de este paso corresponden a toda la información que pudo ser extraída mediante las salidas del paso 1, es decir, toda la respectiva identificación y posterior levantamiento de información con respecto a los datos a gobernar, además de contextualizar la infraestructura tecnológica disponible. Por otro lado, se destaca el trabajo constante con las respectivas áreas implicadas en la arquitectura del dato a gobernar, además de las áreas encargadas de la mantención de las aplicaciones que actualmente gestionan datos dentro de la universidad. En la figura 14 se detallan las respectivas entradas y salidas consideradas para el paso en cuestión.

Para poder entender la arquitectura que involucra a los atributos “Nombre”, “Apellido Materno” y “Apellido Paterno” se tiene que entender el actual modelo de datos que existe en la universidad de Chile, considerando las respectivas bases de datos, y las tablas más pobladas en función de los datos a gobernar en alcance con esta esta memoria.

### 8.3.1 Bases de datos institucionales

Actualmente la universidad cuenta con 2 bases de datos en los que hay existencia de los atributos “Nombre”. “Apellido Materno” y “Apellido Paterno”. Dentro de esta distribución, la base de datos más poblada es “SUG”, asimismo, conteniendo las 2 de las 3 tablas con más registros; “INDIVIDUO” y “DOTACIÓN”.

En primer lugar, la tabla “INDIVIDUO” posee 28 columnas y 4.772.100 registros, donde en esta se almacenan todas las personas que tienen y han tenido algún tipo de relación con la Universidad de Chile, ya sea funcionarios, académicos, alumnos, prestaciones, auxiliares, etc. Esta tabla tiene por propósito poder almacenar los datos personales pertenecientes a estas personas, tales como nombres, sexo, fecha de nacimiento, dirección, entre otros. Actualmente la distribución de atributos que se gobernarán dentro de la tabla “INDIVIDUO” se presenta en la tabla 6.

⚡ PATERNO	⚡ MATERNO	⚡ NOMBRES
Saavedra Nuñez Edna	Jara	NICOLAS
Varela	Oliva	MARIA NATALIA
Carreno	Pena	GABRIELA YAEL
Von	Perez	Joaquín
Ruiz	Jofre	SEBASTIAN
Coletti	Albornoz	IAN LOREDANA
Fierro	Melendez	VICENTE HERNAN
Castro	Astorga	NICOLE ALEJANDRA

*Tabla 6 Ejemplo de registros dentro de la tabla Individuo*

Asimismo, “DOTACIÓN” posee 101 columnas y 112.527 registros, donde se congrega toda la información respectiva de los académicos y funcionarios de la universidad, considerando todos los registros históricos tanto de pregrado como postgrado de todas las facultades y recintos académicos correspondientes a dependencias de la Universidad de Chile. Actualmente, a modo referencial la distribución de atributos que se gobernarán dentro de los alcances de esta memoria dentro de la tabla “DOTACIÓN” se presenta en la tabla 7.

PATERNO	MATERNO	NOMBRES
Saavedra Nuñez Edna	(null)	Juan Carlos
Varela	López	Jaime Robinson
Carreno	Barros	Kathiuska De La Paz
Von	Knorring Cifuentes	Roxana Jacqueline
Ruiz	Coggiola	Magaly Eliana
Coletti	Coletti	Victor Enrique
Fierro	Middleton	Claudio Andres
Castro	Valdes	Zunilda Graciela

Tabla 7 Ejemplo de registros dentro de la tabla Dotación

Por otro lado, en la base de datos WEB, se destaca la tabla “DATOS PERSONALES”, la cual posee 13 columnas y 337.261 registros. Esta tabla contiene en su mayoría información que ingresan los mismos usuarios de la universidad (Estudiantes, funcionarios, y académicos) dentro de aplicaciones que actualmente posee la Universidad de Chile, y considerando que los usuarios que ingresan información poseen una gran libertad de ingreso, debido a la inexistencia de validación en algunas aplicaciones con están conectadas a esta tabla, repercute directamente en la calidad que poseen estos datos. Actualmente la distribución de atributos que se gobernarán dentro de los alcances de esta memoria dentro de la tabla “DATOS PERSONALES” se presenta en la tabla 8.

TIPO_INDIV	MAIL	USERNAME	PATERNO	MATERNO	NOMBRES
Alumno	.uchile.cl	nicolasmeza	Saavedra Nuñez Edna	(null)	Juan Carlos
Alumno	@mail.cl	natigonz	Varela	López	Jaime Robinson
Alumno	@ug.uchile.cl	gabrielasegal	Carreno	Barros	Kathiuska De La Paz
Externo	@mail		Von	Knorring Cifuentes	Roxana Jacqueline
Alumno	@gmail.com	svega	Ruiz	Coggiola	Magaly Eliana
Alumno	sin@mail.cl	i'aempfer	Coletti	Coletti	Victor Enrique
Alumno	@uchile.cl	vlazo	Fierro	Middleton	Claudio Andres
Alumno	@gmail.com	nic@olita	Castro	Valdes	Zunilda Graciela

Tabla 8 Ejemplo de registros dentro de la tabla Datos Personales

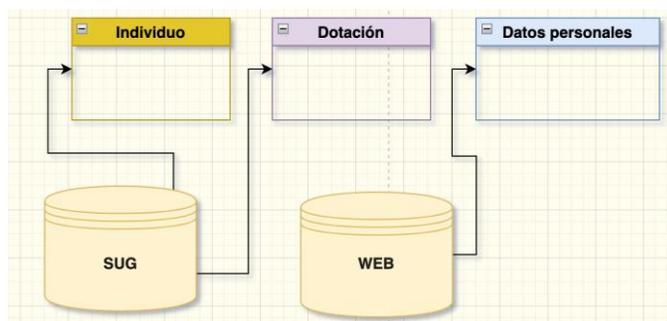


Figura 20 Bases de datos

Para esquematizar la estructura que actualmente se posee con respecto a las bases de datos de la universidad, en la figura 20 se despliega esta, donde las tablas “INDIVIDUO” y “DOTACIÓN” se encuentran dentro de la base de datos “SUG”, mientras que para el caso de la tabla “DATOS PERSONALES”, esta se encuentra dentro del base de datos “WEB”, siendo estas en particular las respectivas tablas a gobernar dentro de los alcances de esta memoria. Esta distinción es importante de mencionar, dado que las herramientas tecnológicas levantadas a posterioridad deben considerar

la respectiva integración entre estas distintas bases de datos. Luego de exponer el esquema actual de base de datos en la universidad, se hace necesario poder entender cómo estos datos están ingresados a estas, y mediante qué procedimientos y aplicaciones.

### 8.3.2 Ingreso de data en aplicaciones

Para efectos de los atributos a gobernar, existen 8 aplicaciones que gestionan el ingreso, modificación y eliminación en las respectivas bases de datos de la universidad. La gran mayoría de las aplicaciones que actualmente se están utilizando para la gestión de datos en la universidad se catalogan como “*legacy applications*”, es decir, aplicaciones que han sido concebidas y desarrolladas con lenguajes o arquitecturas que ya no se emplean en la actualidad, además de presentar paradigmas distintos en cuanto a la importancia de la calidad del dato. Es así como gran cantidad de las aplicaciones que están en uso no poseen ningún tipo de validación de calidad a la entrada dentro de las bases de datos institucionales, además de no presentar ningún tipo de categorización en los respectivos campos dentro de los formularios.

De esta manera, estas “*legacy applications*” se entablan dentro de un paradigma de arquitectura que no considera la actual estructura de calidad de datos que hoy en día se trabaja con respecto a la digitalización y automatización de la gestión y administración, además de la toma de decisiones basada en evidencia empírica. Por otro lado, las áreas técnicas de desarrollo de aplicaciones no han tomado un papel preponderante en modificar estas respectivas aplicaciones, debido a los altos costes económicos y de tiempo asociados, lo que ha implicado en seguir manteniendo estas aplicaciones sin ningún tipo de validación a la entrada de datos hacia las bases de datos institucionales.

### 8.3.3 Salida: Flujos de datos sobre aplicaciones

La primera salida correspondiente a este paso son los reportes asociados al actual flujo de datos en las aplicaciones que modifican información sobre las bases de datos institucionales. El plan de trabajo para esta sección se estructura en función de poder abordar las 8 aplicaciones que transaccionan gran parte de los atributos “Nombre”, “Apellido Materno” y “Apellido Paterno”. Para esto, se generaron mesas de trabajo con las respectivas áreas del negocio implicadas en el ingreso y consumo de estos atributos, además de las áreas de desarrollo y mantención de las aplicaciones, y así levantar información fundamental para comprender el flujo de datos sobre las aplicaciones que hoy en día transaccionan los respectivos datos a gobernar.

Las aplicaciones que en la Universidad de Chile actualmente se encuentran activas en el ingreso de datos asociados a los atributos enunciados en el alcance de esta memoria, son las siguientes: MiUchile, Guía matrícula, SISPER, Títulos y grados, SEPA-VID, AUGE, Honorarios y Bienestar.

El objetivo principal de esta salida es la de poder levantar mapas de información sobre aplicaciones con respecto a los datos “Nombre”, “Apellido Materno” y “Apellido Paterno”, todo esto en función de poder generar un entendimiento de cómo actualmente funcionan estos flujos de datos dentro de las aplicaciones de la Universidad, democratizando así el conocimiento, y contribuyendo de gran manera a generar un perfilamiento de la respectiva información.

Los mapas de información levantados para las respectivas aplicaciones se generan en base a los diagramas de flujo de datos. Los diagramas de flujo de datos (DFD) trazan el flujo de la

información para cualquier proceso o sistema. Emplea símbolos definidos, como rectángulos, círculos y flechas, además de etiquetas de texto breves, para mostrar las entradas y salidas de datos, los puntos de almacenamiento y las rutas entre cada destino (*Lucidchart*). Para el caso concreto de la universidad, se entenderá la siguiente simbología expuesta en la figura 21.



*Figura 21 Simbología mapas de información*

Luego de las reuniones realizadas con las áreas implicadas en el negocio, se bosquejaron los respectivos diagramas de datos para cada aplicación que transacciona los atributos “Nombre”, “Apellido Materno” y “Apellido Paterno”, convergiendo la totalidad de la información disponible. Es así como a modo ejemplo del trabajo realizado para esta sección, se expone uno de los ocho diagramas realizados para las aplicaciones en la figura 22. Es así como se destaca la aplicación MiUchile, la que según la Universidad de Chile “está dirigido a toda la comunidad universitaria, la que tiene por función poder facilitar los trámites administrativos y/u obtención de certificados e ingreso de notas, además de permitir conocer información relevante, ya sea liquidaciones de sueldo o resultados de evaluación del desempeño docente o del personal, entre otros”. En cuanto a los atributos a gobernar dentro de los alcances de esta memoria, esta aplicación despliega en su interfaz gran parte de los nombres y apellidos ingresados en las bases de datos, además de permitir nuevos ingresos de data mediante la misma aplicación, tal como se despliega en la figura 23.



Figura 23 Vista aplicación sobre web MiUchile

Por otro lado, el diagrama realizado para esta aplicación se despliega en la figura 22, donde finalmente se congrega toda la información levantada con las respectivas áreas involucradas, definiendo tablas, acciones, y áreas que tienen permisos CRUD sobre la aplicación.

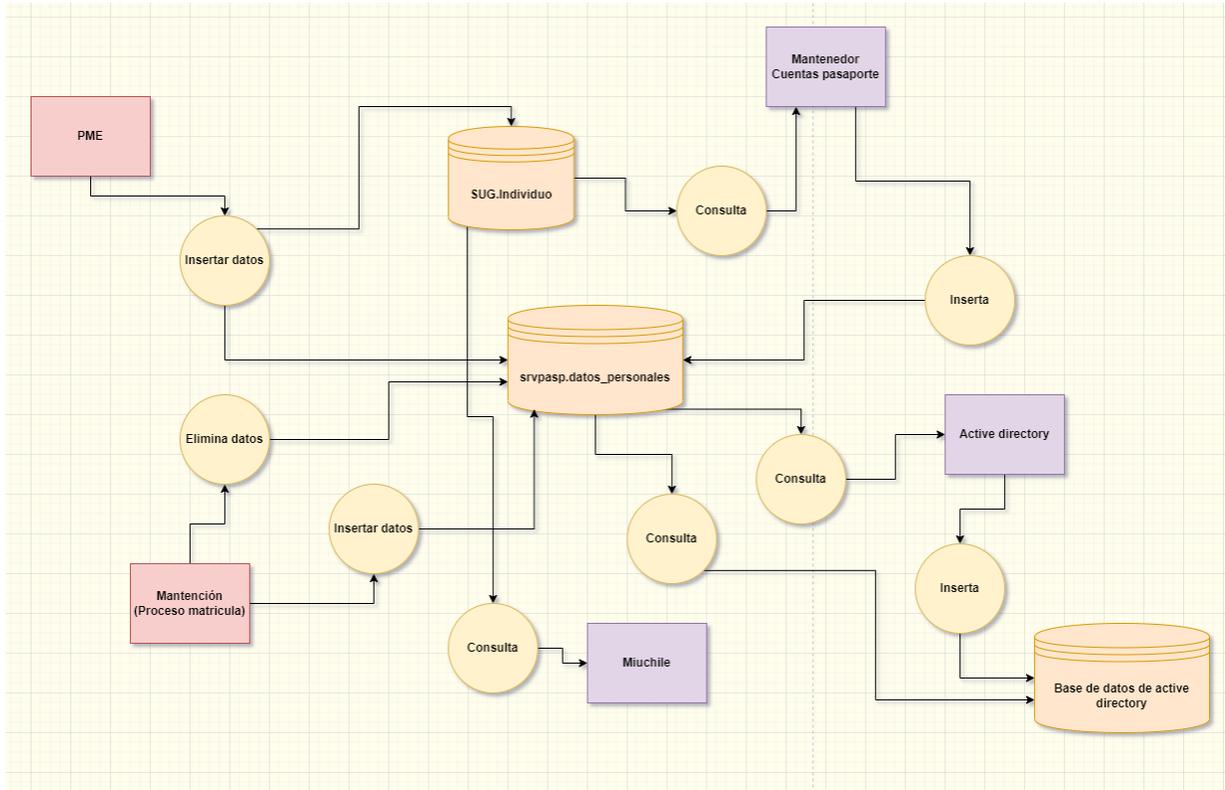


Figura 22 Diagrama de datos MiUchile

Para efectos de los diagramas realizados, en la simbología asociada a los cuadrados, se especifica que los de color morado representan las respectivas aplicaciones que tienen permisos de acción sobre las bases de datos. Por otro lado, los cuadrados rosados corresponden a áreas implicadas en el negocio, que manualmente o mediante las mismas aplicaciones insertan datos en las bases de datos institucionales.

Finalmente, se logra diagramar las 8 aplicaciones que transaccionan los atributos a gobernar, estableciendo flujos de datos importantes para el entendimiento de la infraestructura tecnológica de la universidad, además de comprender de manera específica y transversal el funcionamiento de las aplicaciones en relación con las acciones que respectivamente realizan sobre cada una de las tablas a gobernar. Asimismo, se logra concretar y democratizar esta información a todas las áreas involucradas en la acción sobre los atributos en mención.

### 8.3.4 Salida: Matriz CRUD

Para complementar los diagramas de flujos de datos montados para cada aplicación que transacciona los datos a gobernar, se levanta toda la información sobre permisos de cuentas que explican el funcionamiento de esta aplicación. La función de esta salida es poder identificar los

actuales permisos en operaciones básicas que poseen algunas cuentas sobre las distintas tablas pertenecientes a las bases de datos “SUG” y “WEB”, y en particular a las tablas “INDIVIDUO”, “DOTACIÓN” y “DATOS PERSONALES”, y así dar un complemento importante a los diagramas levantados en la salida anterior.

Esta matriz CRUD fue levantada mediante el servicio de *reporting* que ofrece Microsoft. *Reporting service* propone entregar información correcta a los usuarios del negocio dentro de la Universidad de Chile, permitiendo a estos poder visualizar informes a través de esta plataforma, ya sea por un navegador web, o un teléfono celular. El servicio *reporting service*, permite desarrollar informes y reportes que puedan visualizar distintos entornos a nivel institucional, extrayendo información desde las bases de datos institucionales. Es así como esta herramienta se utiliza para poder desarrollar la respectiva matriz CRUD, nutriéndose de una tabla que posee los correspondientes accesos de los usuarios a las bases de datos, y en particular, a las tablas acordadas para el alcance de esta memoria.

En la figura 23 se presenta la matriz CRUD, la que permite visualizar los respectivos permisos que poseen las cuentas que tienen acceso a las bases transaccionales, pudiendo visualizar gráficamente si es que estas pueden generar acciones de UPDATE, READ, CREATE, o DELETE. Por otro lado, esta matriz también cuenta con 2 filtros, donde el primero de ellos permite realizar una búsqueda del respectivo organismo perteneciente a la Universidad de Chile asociado a la cuenta, y el segundo permite discriminar por tipo de cuenta creada, diferenciando así si pertenecen a cuentas de aplicación, o cuentas de acceso directo a las bases de datos.

Finalmente, considerando las salidas obtenidas para la realización de este paso, se puede levantar información que hasta el momento no existía en la universidad, logrando así identificar tanto la arquitectura del dato en las respectivas aplicaciones de ingreso de datos en la universidad, como los respectivos permisos de acción sobre bases de datos que tienen las distintas cuentas que se le asignan a los usuarios en la actualidad, entendiéndose con más detalle la estructura de funcionamiento sobre la entrada de datos en las bases de datos de la universidad. De esta manera, se da cumplimiento al tercer objetivo específico plasmado para esta memoria, levantando el tratamiento actual en función de los flujos de datos sobre aplicaciones que transaccionan los datos a gobernar.

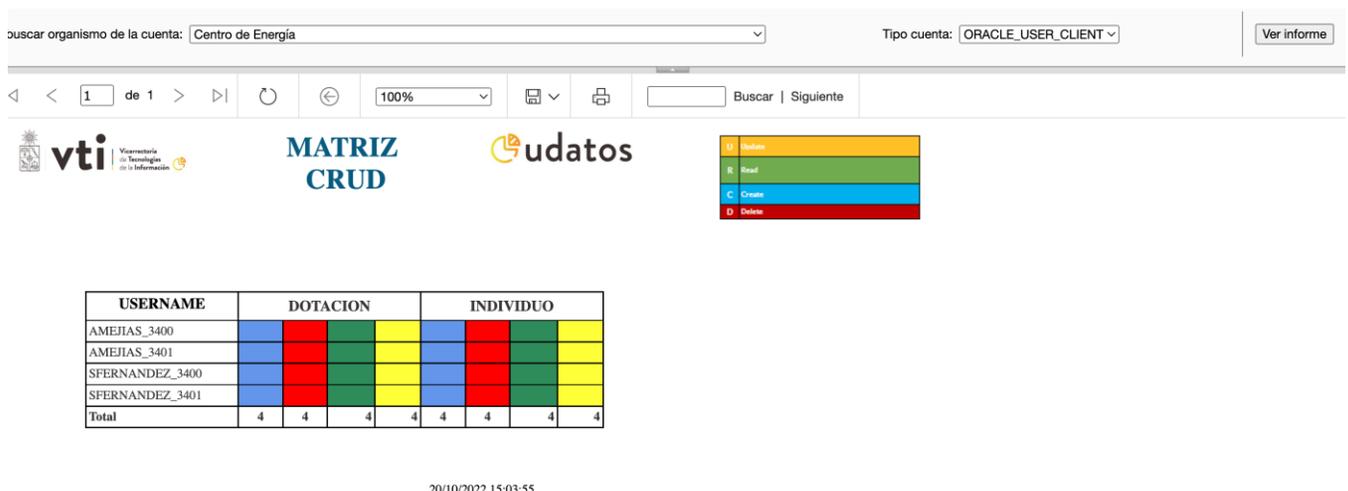


Figura 23 Matriz CRUD

## 8.4 Etapa 2: (Paso 3) Perfilamiento

Luego de poder evaluar los respectivos flujos en las aplicaciones que están operativas hasta el momento, además de la obtención de los permisos de acción sobre las bases de datos transaccionales, se procede a obtener un formato transversal de datos a gobernar a lo largo de todas las unidades de la universidad. Por otro lado, se levantaron y concretaron los indicadores de calidad para poder medir la calidad dentro de las tablas “INDIVIDUO”, “DOTACION”, Y “DATOS PERSONALES”, considerando los atributos a gobernar. En la figura 15 se detallan las respectivas entradas y salidas consideradas para el paso en cuestión.

### 8.4.1 Salida: Levantamiento y definición de indicadores

La primera salida que contempla este paso es la creación de indicadores de calidad que puedan referenciar el estado real y empírico de los datos dentro de las bases de datos institucionales. La labor dentro de esta salida es la de identificar y definir indicadores para la medición de la calidad de datos, y así informar la relación entre los errores de calidad de datos y el no logro de los objetivos propuestos para la gestión institucional.

El DAMA-DBOK ofrece una gran cantidad de indicadores de calidad para la respectiva medición, sin embargo, para efectos de esta memoria, y considerando las dimensiones de los datos a analizar, factibilidad de cálculo, y acceso de información necesaria, fueron elegidos 5 de estos indicadores de calidad a considerar dentro de las bases de datos transaccionales de la universidad, y en específico, las tablas de datos mencionadas previamente. Los indicadores de calidad elegidos fueron exactitud, completitud, unicidad, consistencia y validez.

En función de los indicadores de calidad expuestos y posteriormente seleccionados para realizar las respectivas mediciones, se interpretan para posteriormente ser aplicados los respectivos cálculos sobre cada uno de estos. Finalmente, los cálculos se realizaron de la siguiente manera dentro de las tablas:

- **Completitud:** Se entenderá como el porcentaje de registros con valores no nulos dentro de las respectivas tablas y atributos a analizar.

$$\text{Completitud} = \frac{\sum \text{Valores no nulos}}{\text{Total registros}} \quad (9.1)$$

- **Validez:** Se calculará como el porcentaje de registros que cumplen con los formatos establecidos para la respectiva estandarización, definida en el documento de estandarización de datos para la Universidad de Chile.

$$\text{Validez} = \frac{\sum \text{Valores que cumplen formato}}{\text{Total registros}} \quad (9.2)$$

- Consistencia: Este indicador se calculará como el porcentaje de cruce correcto de una tabla con las demás tablas, entendiendo como cruce correcto la igualdad completa de los registros a comparar.

$$\text{Consistencia} = \frac{\sum \text{Valores que cruzan correctamente con otras tablas}}{\text{Total registros que cruzan}} \quad (9.3)$$

- Exactitud: Mediante registros cargados desde una fuente verídica y confiable, se obtendrá un porcentaje de cruce correcto con los registros que actualmente se obtienen desde las tablas dentro de las bases de datos de la universidad.

$$\text{Exactitud} = \frac{\sum \text{Valores que cruzan correctamente con fuentes real}}{\text{Total registros que cruzan}} \quad (9.4)$$

- Unicidad: Se calculará como el porcentaje de registros que no se repiten o duplican en las tablas analizadas. Para este caso, se contabilizan los valores no duplicados considerando la llave primaria dentro de la respectiva tabla.

$$\text{Unicidad} = \frac{\sum \text{Valores que NO se repiten en llave primaria}}{\text{Total registros dentro de la tabla}} \quad (9.5)$$

Dada esta definición de los cálculos para cada indicador, posteriormente en los próximos pasos se utilizarán para poder generar los respectivos cálculos de la calidad dentro de las tablas y atributos a gobernar. Por otro lado, tomando en consideración la estructuración de los indicadores, el comité estratégico de gobierno de datos, dada la definición alcanzada por los lineamientos presentados por el comité táctico, aprobó el uso de estos, estableciendo requerimientos mínimos de cumplimiento, y de esta manera se establece que los indicadores deben alcanzar un 80% de cumplimiento.

Considerando las salidas obtenidas para la realización de este paso, se logra poder establecer una base importante de definición para las eventuales herramientas tecnológicas que se utilizarán para abordar las problemáticas de los datos a gobernar en cuanto a calidad, además de definir criterios fundamentales de medición y sus respectivas metas.

## 8.4.2 Salida: Estandarización de datos

Luego de identificar los respectivos indicadores que se utilizan para medir la calidad de los datos dentro de las bases de datos, se definen los respectivos formatos, para abordar una estandarización dentro de las tablas “INDIVIDUO”, “DOTACIÓN”. Y “DATOS PERSONALES”. Se entiende como formato a la estructura interna y codificación de los datos en cuestión, permitiendo que estos puedan ser procesados en una forma accesible. En este sentido, es clara la importancia de poder unificar una visión transversal acerca de qué será lo considerado como valido en función de lo presente en las bases de datos, y en específico, las tablas a gobernar.

El establecimiento de estos formatos fue acordado a nivel de comité táctico, con las respectivas áreas implicadas en los atributos en cuestión, presentando de manera transversal los formatos establecidos en las mesas de trabajo, siendo posteriormente aprobadas dentro del comité estratégico para el uso de estos estándares dentro de las bases de datos. Los formatos establecidos son plasmados en la completa guía de uso de formatos para todos los datos personales, además de explicitar ejemplos de uso. El respectivo documento se detalla en el anexo E.

En la tabla 9 se detalla el formato definido para “NOMBRES”, “APELLIDO MATERNO” Y “APELLIDO PATERNO”. El formato definido de manera transversal para la Universidad de Chile es el de considerar la primera letra en mayúsculas, y las demás en minúsculas, tomando en cuenta las respectivas excepciones que se generan en nombres y/o apellidos compuestos.

#### 4.1.2. Nombres

Se registrará en campos separados Nombres, Apellido Paterno y Apellido Materno, donde en cada campo deberá tener registrado la primera letra de la palabra en mayúscula y el resto en minúscula. Para comparación o búsqueda de nombres se debe comparar utilizando el mismo formato, con la finalidad de retornar correctamente la información.

A continuación se detalla un ejemplo:

ID_PERSONA	NOMBRES	PATERNO	MATERNO	GENERO
1	58 Alfredo	Pérez	Herrera	MASCULINO
2	59 Oscar Lizardo	Fernández	Cabezas	MASCULINO
3	60 Rafael	Muñoz	de la Torre	MASCULINO
4	61 Guillermo	Alveal	Rue	MASCULINO
5	62 Felix	Jara	Turkenich	MASCULINO

Tabla 9 Formato definido para Nombres y Apellidos

Finalmente se da cumplimiento al cuarto objetivo específico pactado para esta memoria, logrando perfilar los datos a gobernar, y específicamente, logrando levantar los formatos a utilizar transversalmente dentro de la universidad para los atributos en cuestión, además de definir los respectivos indicadores de calidad que tendrán por objetivo medir la calidad de estos atributos dentro de las tablas a gobernar.

## 8.5 Etapa 2: (Paso 4) Despliegue

El despliegue corresponde al último paso que proyecta esta memoria, y es donde finalmente se congrega toda la información levantada con las distintas áreas involucradas en los datos a gobernar, para así generar soluciones tecnológicas que converjan en una solución factible en términos de monitoreo y limpieza. De esta manera, el presente paso de despliegue se estructura en dos etapas fundamentales para su implementación; monitoreo y limpieza, tal como se detalla en la figura 24.

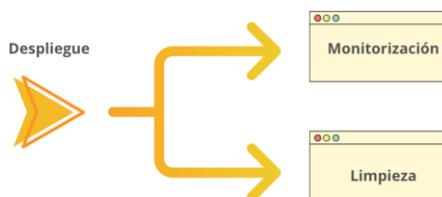


Figura 24 Etapas despliegue

En primer lugar, la etapa de monitorización está enfocada en poder levantar herramientas de monitorización sobre las respectivas tablas y atributos a gobernar dentro de los alcances de la presente memoria, y en específico, poder confeccionar un *dashboard* de calidad que integre los indicadores de calidad levantados en la etapa de perfilamiento, y sus respectivos cálculos dentro de las bases transaccionales de la Universidad de Chile. Por otro lado, la segunda etapa que proyecta este paso es la de poder generar acción de reparación sobre los indicadores que se encuentren bajo los criterios mínimos exigidos por el comité estratégico de gobierno de datos. De esta manera, se integran diferentes herramientas tecnológicas para hacer frente a esta situación tanto de una perspectiva de corto como a largo plazo.

Con respecto a las entradas de este paso, es importante considerar toda la documentación que existe y que ha sido levantada por las distintas mesas para congregar información de los datos a gobernar. Toda esta información permite poder entender a cabalidad las distintas implicancias en el dónde medir, y cómo medir los indicadores de calidad, además de definir los lineamientos de creación del respectivo *dashboard* de calidad, para posteriormente generar una eventual limpieza dentro de los datos.

Por otro lado, se considera importante poder tomar en cuenta todas las herramientas que puedan ser de utilidad a la hora de poder generar soluciones tecnológicas de monitorización y limpieza dentro de los datos. Para este caso, se considera toda la estructura que implica tanto en bases SQL Server y Bases Oracle, además de las herramientas que generen utilidad en función de poder obtener todas las capas de información requeridas para la implementación de estas soluciones. Finalmente, se integran todas las practicas efectivas en inteligencia de negocios para poder converger de manera correcta los distintos indicadores de calidad, y posteriormente poder visualizarlos de una manera eficiente considerando estrategias de monitorización que se definan en este paso.

### 8.5.1 Salida: Creación *Dashboard* de calidad (Monitorización)

Dentro de las mesas de trabajo realizadas a lo largo del desarrollo de esta memoria, una de las necesidades más demandadas dentro de la gestión de datos es poder monitorizar efectivamente como los datos se encuentran actualmente en las respectivas bases de datos de la Universidad de Chile. Para esto, una de las soluciones levantadas para abordar este problema fue la creación de un *dashboard* de calidad, el que permita conocer el estado actual de los datos a gobernar dentro de las respectivas tablas a estudiar, integrando los indicadores de calidad definidos en la etapa de perfilamiento. Una característica fundamental del *dashboard* de calidad proyectado es la de poder conocer la temporalidad de los respectivos indicadores definidos e implementarlos dentro de este, permitiendo al usuario entender los movimientos y fluctuaciones que puedan tener estos indicadores a lo largo del tiempo.

Para poder desarrollar este *dashboard* de calidad, se genera todo un proceso de creación para este, considerando soluciones tecnológicas enfocadas en poder inicialmente calcular los indicadores de calidad, generar migraciones de tablas entres distintas bases de datos, y finalmente poder poblar de datos un panel montado en PowerBI.

Tal como se mencionaba, en primer lugar, se toman los indicadores de exactitud, consistencia, validez, unicidad, y completitud definidos en el paso de perfilamiento, para posteriormente integrarlos a lenguaje SQL en forma de cálculo para cada uno de estos. Por otro lado, también se

considera el cálculo de estos indicadores sobre los respectivos atributos a gobernar (Nombres y apellidos), y las tablas que almacenan estos (“INDIVIDUO”, “DOTACION”, “DATOS PERSONALES”).

A modo excepción, y en función de la naturaleza del indicador de exactitud, el proceso de cálculo para este se da en función de una fuente real de información previamente validada, y de esta manera, se establece una conexión a la API del registro civil, la cual provee gran cantidad de atributos asociados a datos personales, y en específico, los asociados a nombres y apellidos. Posteriormente, para el cálculo del indicador de exactitud se realiza una comparativa con respecto a los datos obtenidos desde el Registro Civil.

Dentro de esta conexión, existe todo un proceso para extraer los datos del registro civil, y finalmente cargarlos en las bases institucionales de la universidad. De esta manera, para comunicarse efectivamente con la API del registro civil, se establece un proceso que presenta una estructura de autorización mediante *tokens*, en donde el usuario realiza peticiones que son procesadas mediante permisos aplicados dentro de la API, y las respectivas bases del Registro Civil. El usuario genera estas peticiones mediante un script realizado en JavaScript, el cual está soportado por el entorno de ejecución Node.js, que permite a su vez comunicarse con la API del registro civil, obtener los respectivos outputs en formato JSON, y finalmente poder integrar estos datos dentro de las bases de datos transaccionales de la Universidad de Chile

Dentro del proceso implicado para comunicarse con la API del Registro Civil, existe una serie de pasos para cargar efectivamente los datos dentro de las bases transaccionales de la universidad, los que se detallan en el anexo F. El proceso parte con el ingreso de credenciales de entrada y el RUT a consultar por parte del usuario dentro del script realizado en el módulo NODE.js (El script es expuesto en el anexo G, a modo referencia). El script ingresa las respectivas credenciales, autenticándose en la API que ofrece el registro civil, para luego recibir un *token* de acceso, permitiendo al script poder entregar este *token* de acceso y el RUT a consultar a la API del registro civil, donde se retorna el respectivo JSON con los datos personales asociados al RUT, y finalmente el script integra estos *outputs* dentro de las bases de datos transaccionales de la universidad, en una tabla de naturaleza *STAGE*. Luego de obtener estos resultados en la tabla *STAGE*, se procede a generar el respectivo cálculo de este indicador mediante consultas SQL. En términos simples, la API del registro civil provee datos, para posteriormente ser almacenados en las bases de datos transaccionales de la universidad, y generar los respectivos cálculos para el indicador de exactitud.

Es así como luego de generar estos cálculos para exactitud y los demás indicadores, se da inicio a todo un proceso de poblamiento de datos hacia el *dashboard* de calidad, desde las bases transaccionales de la universidad. Este proceso congrega la creación de tablas temporales dentro de las bases de datos para almacenar estos cálculos, y luego ser procesados de la manera correcta, para finalmente integrarlos al *dashboard* de calidad. El proceso se detalla mediante el siguiente BPMN descrito en el anexo H.

El proceso expuesto en el anexo H fue montado dentro de la base de datos de datos “SUG”, alojado en servidores ORACLE, y un *datamart* alojado en el servidor SQL Server, los que, mediante una cuenta con permisos otorgados, posee una integración hacia otras bases de datos. Esta integración se realiza mediante un DBLINK, el cual es un comando SQL que permite poder comunicar tablas provenientes de distintas bases de datos.

El proceso parte con la creación de estas consultas, considerando todos los indicadores de calidad enunciados para efectos de esta memoria, los que son congregados dentro de un procedimiento almacenado, el que tiene por función ingresar a una tabla temporal todo el cálculo de los respectivos indicadores, considerando los distintos atributos y fuentes. A modo referencial, el respectivo procedimiento almacenado se muestra en el anexo I. Todos estos cálculos son llevados a la tabla temporal que se genera para este proceso, en donde los respectivos campos que se generaron son los indicados en la tabla 10.

NO_CUMPLE	CUMPLE	TOTAL	ATRIBUTO	FUENTE	INDICADOR	ORIGEN
1	228	2794	3022 NOMBRES	INDIVIDUO	EXACTITUD	REGISTRO CIVIL
2	136	2886	3022 PATERNO	INDIVIDUO	EXACTITUD	REGISTRO CIVIL
3	141	2871	3012 MATERNO	INDIVIDUO	EXACTITUD	REGISTRO CIVIL
4	36	2986	3022 FECHA_NACIMIENTO	INDIVIDUO	EXACTITUD	REGISTRO CIVIL
5	14	3008	3022 SEXO	INDIVIDUO	EXACTITUD	REGISTRO CIVIL
6	35567	147554	183121 CIUDAD	INDIVIDUO	EXACTITUD	REGISTRO CIVIL
7	94678	26684	121362 PAIS	INDIVIDUO	EXACTITUD	REGISTRO CIVIL
8	3404	15168	18572 NOMBRES	DOTACION	EXACTITUD	REGISTRO CIVIL
9	2374	16272	18646 PATERNO	DOTACION	EXACTITUD	REGISTRO CIVIL

*Tabla 10 Tabla temporal con cálculo de indicadores*

Luego de que el procedimiento almacenado compile correctamente, y se integre todos los valores asociados a indicadores dentro de la tabla temporal, se configura un JOB periódico, el cual tiene como función poder gatillar este procedimiento almacenado diariamente, y así poder obtener la temporalidad de los indicadores de calidad. En otras palabras, este procedimiento almacenado es configurado para que genere cálculo de indicadores todos los días, y así dar seguimiento a estos. Al tener los respectivos resultados de los cálculos de los indicadores para todos los atributos y fuentes, estos entran a un ETL desarrollado en Visual Studio, el cual posee la siguiente estructura detallada en el anexo J, teniendo por función poder migrar datos desde distintos gestores de bases de datos (Oracle a SQL Server), y finalmente poblar el panel de calidad.

El respectivo BPMN que detalla el ETL se expone en el anexo K. Para este proceso se empieza con una limpieza de la tabla temporal que está dentro del base Oracle, la que se crea para limpiar cualquier registro que se haya cargado con antelación a la fecha presente de carga. Tras la limpieza, se migran los datos desde esta tabla establecida en las bases Oracle hacía la base SQL Server. Luego de migrar, se limpian los datos de la tabla temporal dentro de la base SQL Server, para así evitar más de una carga por día, y no obtener duplicados de datos. Finalmente, estos datos dentro de la tabla temporal son cargados dentro de una tabla normal, la que se va poblando en función de las cargas de las fechas anteriores, y así poder tener el registro temporal de los respectivos indicadores. En términos simples, esta etapa del proceso migra datos desde las bases Oracle a SQL server, y se asegura de que no exista duplicados en las tablas creadas.

Luego de obtener los datos de la tabla final en SQL Server, se procede a generar el *dashboard* de calidad en PowerBI, el que congrega un análisis tanto general para las respectivas tablas, como un análisis más detallado con respecto a los atributos que se gobernarán. Dentro del *dashboard* de calidad expuesto ya se pueden obtener los respectivos cálculos que posee cada indicador de calidad para cada tabla considerando los atributos en cuestión, además de incluir una vista asociada a la temporalidad de estos indicadores.

Es así como finalmente se integra una solución de monitorización que permite a los usuarios y gestores de calidad de datos poder dar seguimiento a los indicadores de calidad dentro de las tablas

a gobernar, y servir como referencia para eventuales acciones correctivas a realizar dentro de los atributos y tablas a gobernar.

## 8.5.2 Resultados Monitorización

El *dashboard* propuesto integra diferentes vistas para visualizar los indicadores de calidad dentro de estas tablas y atributos. En primer lugar, se presenta una visualización general de las tablas a gobernar en función de los indicadores calculados, considerando los promedios que se obtienen para los atributos nombres y apellidos. Asimismo, existe una pestaña asociada a obtener un análisis por identificador (RUT) dentro de las tablas y atributos a gobernar, es decir, poder obtener los registros que cumplen correctamente con los indicadores de calidad, considerando tanto nombres como apellidos, por ejemplo, si es que para un RUT específico dentro de algunas de las tablas a gobernar, alguno los atributos “Nombre”, “Apellido Materno” y “Apellido Paterno” no cumple con el indicador a estudiar, el identificador completo no cumple con el respectivo indicador, y de esta manera, se genera un análisis más detallado y sin sesgos. Por otro lado, existe una visualización que permite obtener los resultados por tablas a gobernar, pero desplegando un filtro para los respectivos atributos, generando así un análisis más detallado para los registros a gobernar, en particular “Nombre”, “Apellido Materno” y “Apellido Paterno”.

En las siguientes subsecciones se expondrán las distintas visualizaciones del *dashboard* de calidad, además de desplegar los resultados obtenidos, para cada una de estas.

### 8.5.2.1 *Dashboard* de calidad sobre tablas (Nombres y Apellidos)

En la figura 25, 26, y 27 se expone el respectivo *dashboard* de calidad generado para las tablas “INDIVIDUO”, “DOTACION”, y “DATOS PERSONALES”, considerando los atributos definidos dentro de los alcances, en específico, “Nombres”, “Apellido Materno”, y “Apellido Paterno”. Esta caracterización de indicadores en función de las tablas a gobernar nos permite comprender cómo actualmente se encuentra el promedio de cálculo de indicadores de los atributos asociados a nombres y apellidos sobre estas. En otras palabras, esta gráfica expone los resultados de calcular el porcentaje de cumplimiento de los indicadores para cada atributo por separado dentro de una tabla específica, y después, para este mismo indicador, calcular el promedio de lo obtenido para cada atributo, donde a modo ejemplo, se toma el siguiente caso hipotético; tomando un indicador específico, y una tabla específica, se obtiene un  $a\%$  para “Nombre”, un  $b\%$  para “Apellido Materno”, y un  $c\%$  para “Apellido Paterno”, y por tanto, el resultado final para este indicador sobre esta tabla será el promedio entre  $a$ ,  $b$  y  $c$ , esto considerando que para cada uno de los cálculos de los porcentaje para los atributos, se posee la misma cantidad de registros como total.

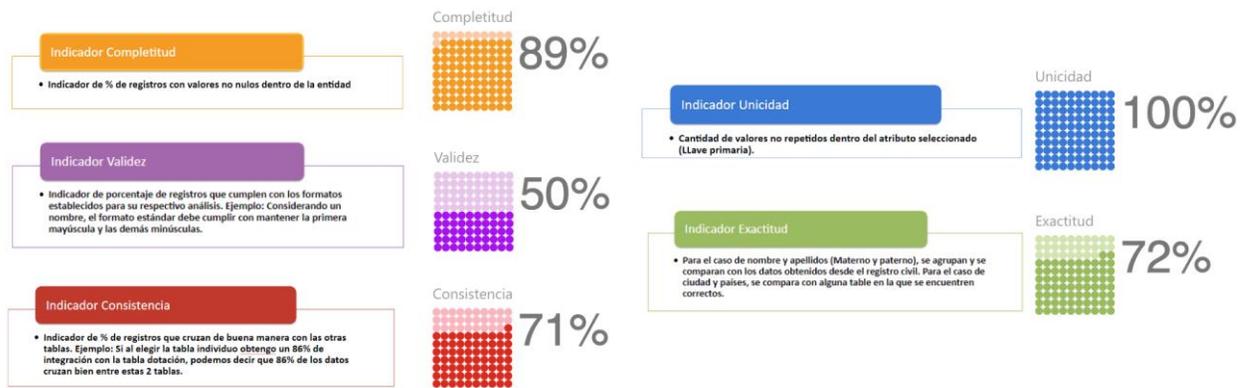


Figura 25 Dashboard de calidad (Individuo)

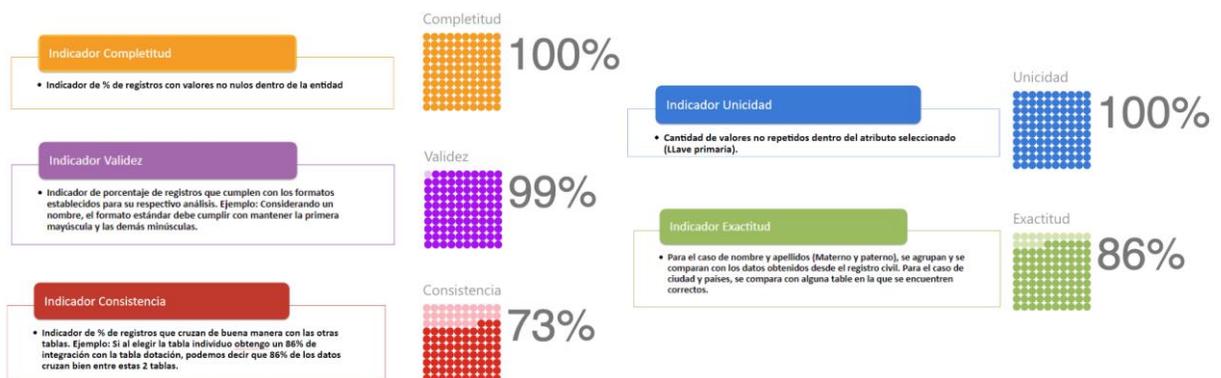


Figura 26 Dashboard de calidad (Dotación)

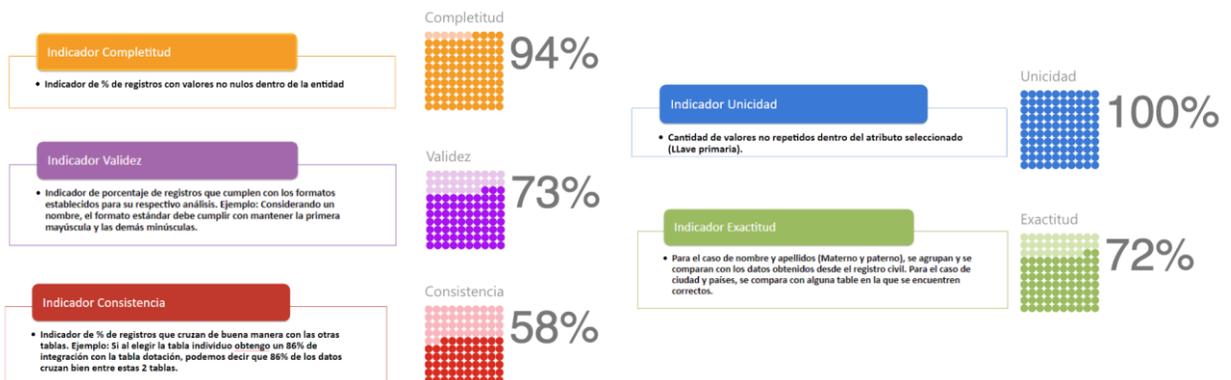


Figura 27 Dashboard de calidad (Datos Personales)

En primer lugar, con respecto a la figura 25, y en específico, al cálculo de los indicadores sobre la tabla “INDIVIDUO”, se puede desprender los siguientes resultados; completitud un 89%, validez un 50%, consistencia un 71%, unicidad un 100%, y finalmente un 72% de exactitud. Los porcentajes de cumplimiento para los indicadores de validez, consistencia, y exactitud exponen resultados por bajo los requerimientos mínimos aprobados por el comité estratégico de gobierno de datos.

Con respecto a la figura 26, y en específico, al cálculo de los indicadores sobre la tabla “DOTACION”. Los resultados obtenidos fueron los siguientes; completitud un 100%, validez un

99%, consistencia un 73%, unicidad un 100%, y finalmente un 86% de exactitud. En cuanto al porcentaje de cumplimiento para el indicador de consistencia, expone resultados por bajo los requerimientos mínimos aprobados por el comité estratégico de gobierno de datos.

Finalmente, en la figura 27, y en específico, al cálculo de los indicadores sobre la tabla “DATOS PERSONALES”, se puede desprender los siguientes resultados; completitud un 94%, validez un 73%, consistencia un 58%, unicidad un 100%, y finalmente un 72% de exactitud. En términos generales, los porcentajes de cumplimiento para los indicadores de validez, consistencia, y exactitud exponen resultados por bajo los requerimientos mínimos aprobados por el comité estratégico de gobierno de datos.

### 8.5.2.2 *Dashboard* de calidad sobre tablas (Nombres y Apellidos por identificador)

Dado que los resultados obtenidos para la visualización anterior estarían considerando el promedio de los atributos “Nombre”, “Apellido Materno” y “Apellido Paterno”, para un respectivo indicador, se estaría sesgando el análisis, dado que en un caso hipotético; si se toma un indicador específico, y se obtiene un  $x\%$  para “Nombre”, un  $y\%$  para “Apellido Materno”, y un  $z\%$  para “Apellido Paterno”, donde  $x \gg y \gg z$ , el resultado del promedio estaría subestimando el impacto que actualmente tendría el  $x\%$  del atributo “Nombre”. De esta manera, se propone este análisis por identificador (RUT), el que despliega un nuevo análisis uniendo estos subconjuntos relacionados a nombres y apellidos en un nuevo cálculo para los respectivos indicadores definidos anteriormente, considerando los resultados por identificador de personas (RUT), utilizando las nuevas fórmulas expresadas a continuación:

$$\text{Completitud} = \frac{\sum \text{Valores no nulos tanto en nombres como apellido por identificador}}{\text{Total registros por identificador}} \quad (9.6)$$

$$\text{Validez} = \frac{\sum \text{Valores por identificador que cumplen formato tanto en nombres como apellidos}}{\text{Total registros por identificador}} \quad (9.7)$$

$$\text{Consistencia} = \frac{\sum \text{Valores que cruzan correctamente tanto en nombres como apellidos con otras tablas}}{\text{Total registros que cruzan por identificador}} \quad (9.8)$$

$$\text{Exactitud} = \frac{\sum \text{Valores por identificador que cruzan correctamente tanto en nombres como apellidos con fuente real}}{\text{Total registros que cruzan por identificador}} \quad (9.9)$$

$$\text{Unicidad} = \frac{\sum \text{Valores que no se repiten por identificador tanto en nombres y apellidos}}{\text{Total registros dentro de la tabla por identificador}} \quad (9.10)$$

De esta manera, se consigue obtener la cantidad de atributos por identificador (RUT), que dan cumplimiento a los indicadores de calidad considerando tanto sus respectivos nombres como apellidos. En la figura 28, 29, y 30 se expone el respectivo *dashboard* de calidad generado para las tablas “INDIVIDUO”, “DOTACION”, y “DATOS PERSONALES”, considerando el análisis de los atributos “Nombres”, “Apellido Materno”, y “Apellido Paterno” por identificador.

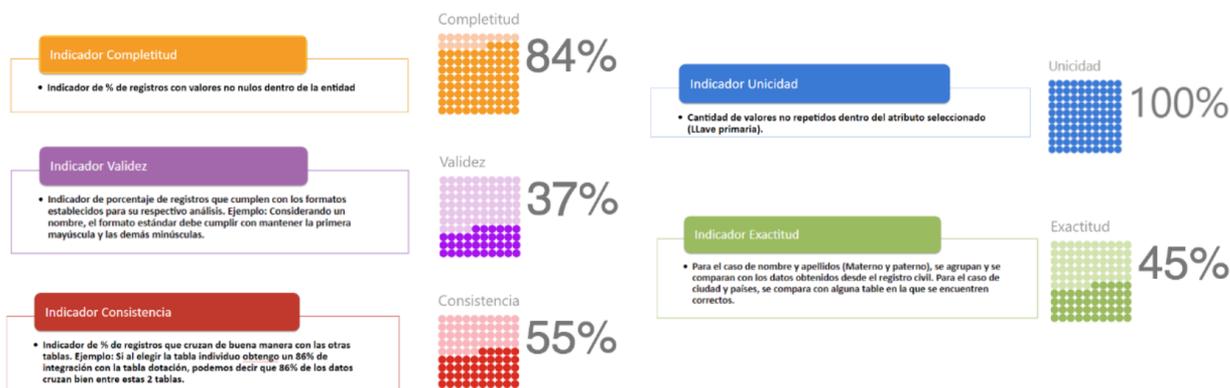


Figura 28 Dashboard de calidad (Individuo)



Figura 29 Dashboard de calidad (Dotación)



Figura 30 Dashboard de calidad (Datos Personales)

En primer lugar, con respecto a la figura 28, y en específico, al cálculo de los indicadores sobre la tabla “INDIVIDUO”, se puede desprender los siguientes resultados; completitud un 84%, validez un 37%, consistencia un 55%, unicidad un 100%, y finalmente un 45% de exactitud. Los porcentajes de cumplimiento para los indicadores de validez, consistencia, y exactitud exponen

resultados por bajo los requerimientos mínimos aprobados por el comité estratégico de gobierno de datos.

Con respecto a la figura 29, y en específico, al cálculo de los indicadores sobre la tabla “DOTACION”, los resultados obtenidos fueron los siguientes; completitud un 100%, validez un 86%, consistencia un 64%, unicidad un 100%, y finalmente un 58% de exactitud. En cuanto al porcentaje de cumplimiento para los indicadores de consistencia y exactitud, exponen resultados por bajo los requerimientos mínimos aprobados por el comité estratégico de gobierno de datos.

Finalmente, en la figura 30, y en específico, al cálculo de los indicadores sobre la tabla “DATOS PERSONALES”, se puede desprender los siguientes resultados; completitud un 83%, validez un 66%, consistencia un 51%, unicidad un 100%, y finalmente un 37% de exactitud. En términos generales, los porcentajes de cumplimiento para los indicadores de validez, consistencia, y exactitud exponen resultados por bajo los requerimientos mínimos aprobados por el comité estratégico de gobierno de datos

### 8.5.2.3 Dashboard de calidad sobre atributo “Nombre”

Por otro lado, en la figura 31, 32 y 33 se despliega la visualización del atributo “Nombre” dentro de las 3 tablas a gobernar; “INDIVIDUO”, “DOTACION”, y “DATOS PERSONALES”. Esta visualización permite poder integrar una visión particular para el atributo “Nombre” en función de los indicadores de calidad definidos.

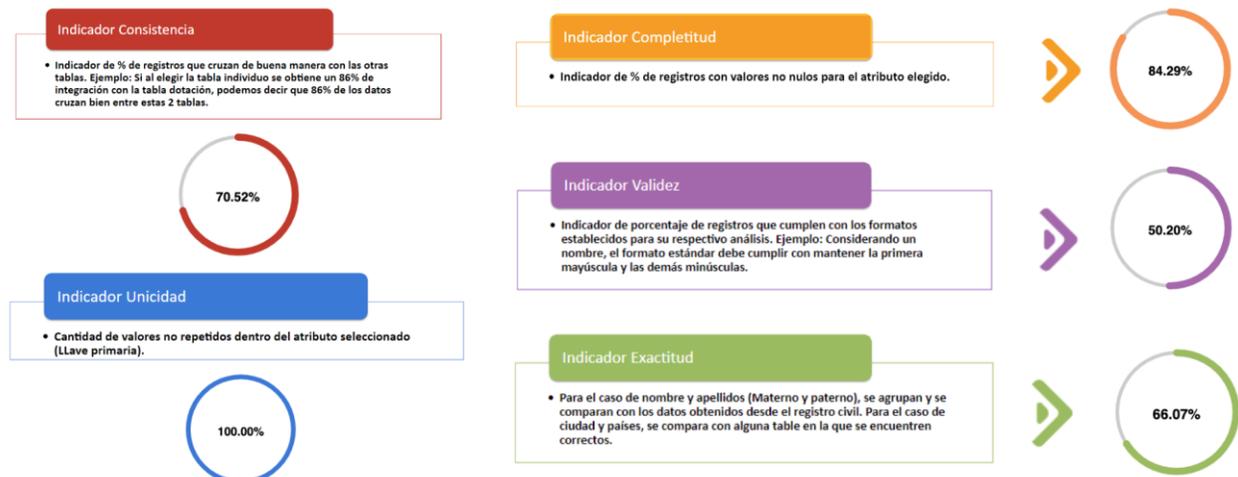


Figura 31 Dashboard de calidad (Nombres en Individuo)



Figura 32 Dashboard de calidad (Nombres en Dotación)

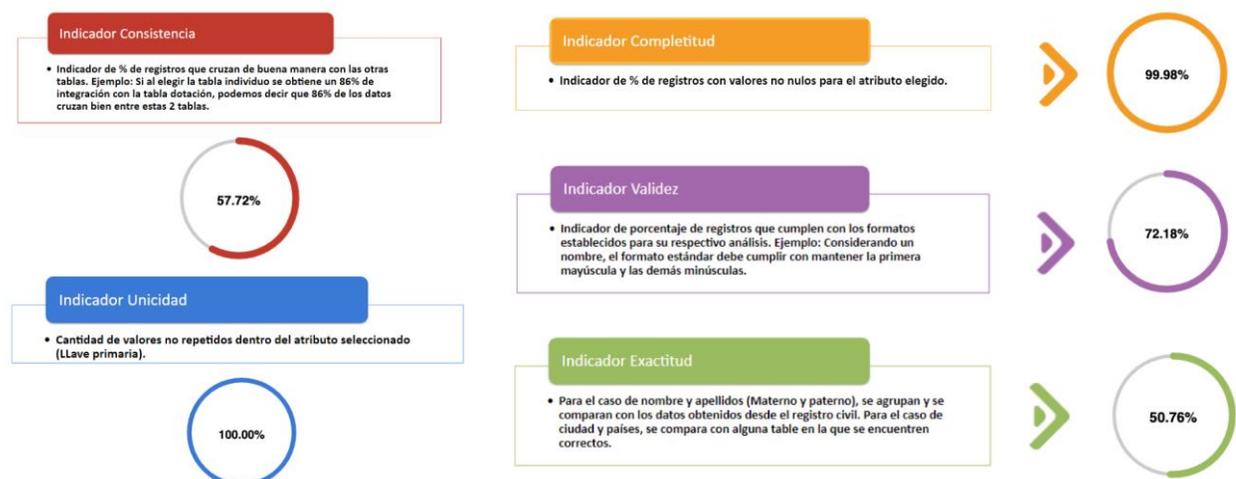


Figura 33 Dashboard de calidad (Nombres en Datos Personales)

Con respecto a la figura 31, y en específico, al cálculo de los indicadores asociados al atributo “Nombre” sobre la tabla “INDIVIDUO”, se obtienen los siguientes resultados; completitud un 84%, validez un 50%, consistencia un 71%, unicidad un 100%, y finalmente un 66% de exactitud. En términos generales, los porcentajes de cumplimiento de tres indicadores exponen resultados por bajo los requerimientos mínimos aprobados por el comité estratégico de gobierno de datos, en específico los indicadores de validez, consistencia, y exactitud.

Por otro lado, en la figura 32, y en específico, al cálculo de los indicadores asociados al atributo “Nombre” sobre la tabla “DOTACION”, los resultados obtenidos fueron los siguientes; completitud un 100%, validez un 99%, consistencia un 73%, unicidad un 100%, y finalmente un 77% de exactitud. En términos generales, los porcentajes de cumplimiento de dos indicadores exponen resultados fuera de los requerimientos mínimos aprobados por el comité estratégico de gobierno de datos, en específico los indicadores de consistencia y exactitud.

Finalmente, en la figura 33, y en específico, al cálculo de los indicadores asociados al atributo “Nombre” sobre la tabla “DATOS PERSONALES”, se obtienen los siguientes resultados; completitud un 100%, validez un 72%, consistencia un 58%, unicidad un 100%, y finalmente un

51% de exactitud. En términos generales, los porcentajes de cumplimiento para tres indicadores exponen resultados fuera de los requerimientos mínimos aprobados por el comité estratégico de gobierno de datos, en específico los indicadores de consistencia, exactitud y validez.

### 8.5.2.4 Dashboard de calidad sobre atributo “Apellido materno”

Siguiendo la misma estructura anterior, en la figura 34, 35 y 36 se despliega la visualización del atributo “Apellido Materno” dentro de las 3 tablas a gobernar; “INDIVIDUO”, “DOTACION”, y “DATOS PERSONALES”. Esta visualización permite poder integrar una visión particular para el atributo “Apellido Materno” en función de los indicadores de calidad definidos.

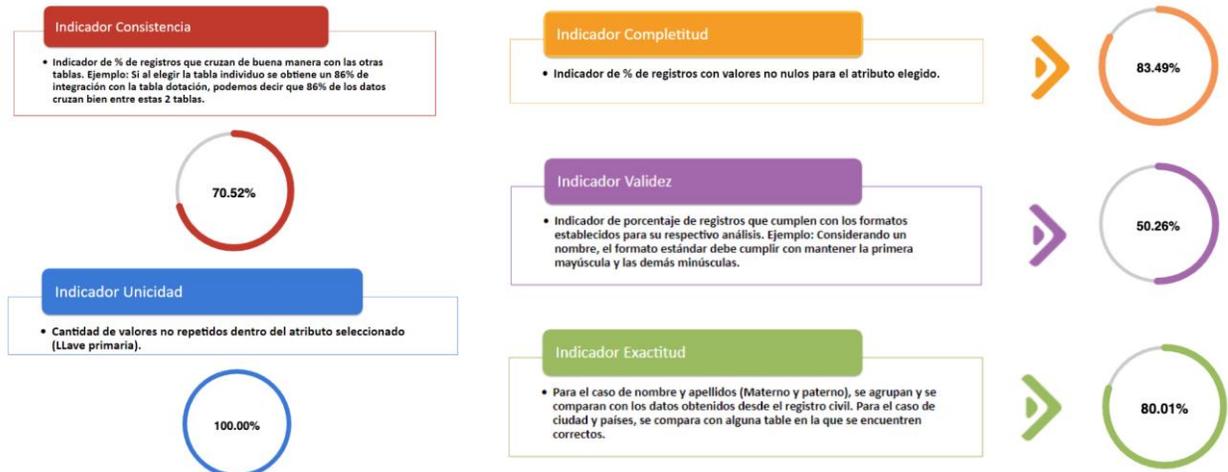


Figura 34 Dashboard de calidad (Apellido Materno en Individuo)

Con respecto a la figura 34, y en específico, al cálculo de los indicadores asociados al atributo “Apellido Materno” sobre la tabla “INDIVIDUO”, se obtienen los siguientes resultados; completitud un 83%, validez un 50%, consistencia un 71%, unicidad un 100%, y finalmente un 80% de exactitud. En términos generales, los porcentajes de cumplimiento para dos indicadores obtenidos exponen resultados por bajo los requerimientos mínimos aprobados por el comité estratégico de gobierno de datos, en específico los indicadores de validez y consistencia.



Figura 35 Dashboard de calidad (Apellido Materno en Dotación)

Con respecto a la figura 35, y en específico, al cálculo de los indicadores asociados al atributo “Apellido Materno” sobre la tabla “DOTACION”, se obtienen los siguientes resultados; completitud un 100%, validez un 100%, consistencia un 73%, unicidad un 100%, y finalmente un 99% de exactitud. El porcentaje de cumplimiento para el indicador de consistencia expone resultados bajo los requerimientos mínimos aprobados por el comité estratégico de gobierno de datos.



Figura 36 Dashboard de calidad (Apellido Materno en Datos Personales)

Con respecto a la figura 36, y en específico, al cálculo de los indicadores asociados al atributo “Apellido Materno” sobre la tabla “DATOS PERSONALES”, se obtienen los siguientes resultados; completitud un 83%, validez un 75%, consistencia un 58%, unicidad un 100%, y finalmente un 83% de exactitud. En términos generales, los porcentajes de cumplimiento para dos indicadores obtenidos exponen resultados por bajo los requerimientos mínimos aprobados por el comité estratégico de gobierno de datos, en específico los indicadores de consistencia, y validez.

#### 8.5.2.5 Dashboard de calidad sobre atributo “Apellido paterno”

Finalizando el análisis de los atributos a gobernar dentro de los alcances de esta memoria, en la figura 37, 38 y 39 se despliega la visualización del atributo “Apellido Paterno” dentro de las 3 tablas a gobernar; “INDIVIDUO”, “DOTACION”, y “DATOS PERSONALES”. Esta visualización permite poder integrar una visión particular para el atributo “Apellido Paterno” en función de los indicadores de calidad definidos.

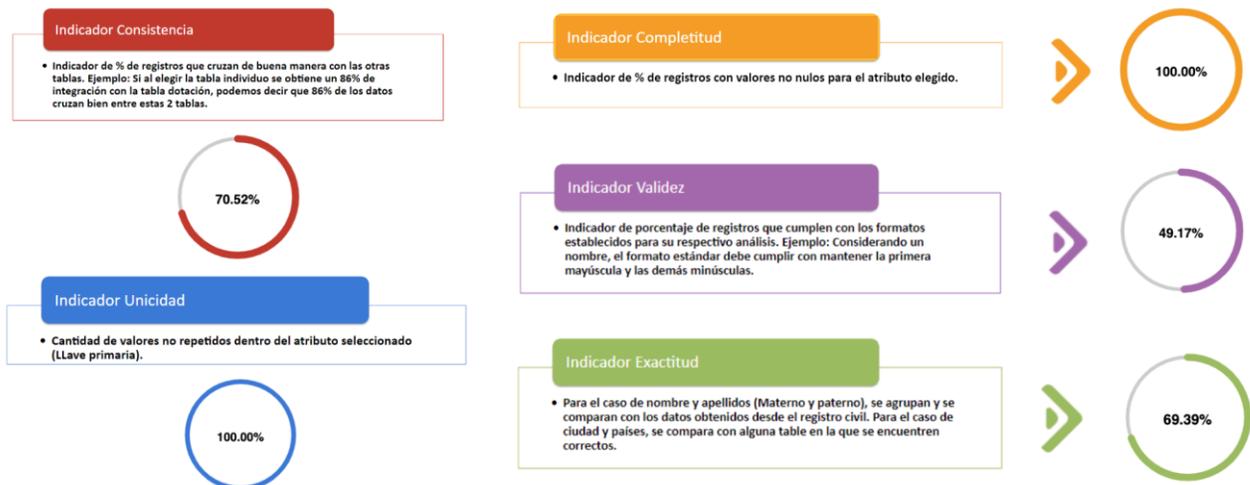


Figura 37 Dashboard de calidad (Apellido Paterno en Individuo)

Con respecto a la figura 37, y en específico, al cálculo de los indicadores asociados al atributo “Apellido Paterno” sobre la tabla “INDIVIDUO”, se obtienen los siguientes resultados; completitud un 100%, validez un 49%, consistencia un 71%, unicidad un 100%, y finalmente un 69% de exactitud. En términos generales, los porcentajes de cumplimiento para tres indicadores obtenidos exponen resultados por bajo los requerimientos mínimos aprobados por el comité estratégico de gobierno de datos, en específico los indicadores de validez, exactitud y consistencia.



Figura 38 Dashboard de calidad (Apellido Paterno en Dotación)

Con respecto a la figura 38, y en específico, al cálculo de los indicadores asociados al atributo “Apellido Paterno” sobre la tabla “DOTACION”, se obtienen los siguientes resultados; completitud un 100%, validez un 100%, consistencia un 73%, unicidad un 100%, y finalmente un 82% de exactitud. En términos generales, el porcentaje de cumplimiento para un indicador obtenido expone resultados por bajo los requerimientos mínimos aprobados por el comité estratégico de gobierno de datos, en específico el indicador de consistencia.



Figura 39 Dashboard de calidad (Apellido Paterno en Datos Personales)

Con respecto a la figura 39, y en específico, al cálculo de los indicadores asociados al atributo “Apellido Paterno” sobre la tabla “DATOS PERSONALES”, se obtienen los siguientes resultados; completitud un 100%, validez un 73%, consistencia un 58%, unicidad un 100%, y finalmente un 81% de exactitud. En términos generales, los porcentajes de cumplimiento para dos indicadores obtenidos exponen resultados por bajo los requerimientos mínimos aprobados por el comité estratégico de gobierno de datos, en específico los indicadores de validez y consistencia.

### 8.5.3 Resumen resultados obtenidos

Luego de implementar las respectivas vistas en el *dashboard* de calidad con los cálculos de los indicadores, se pueden ver resultados parciales en algunos indicadores de calidad bajo los criterios exigidos por el comité estratégico de gobierno de datos. Dentro de los cálculos realizados para las tablas, considerando el promedio de los indicadores, desplegados en las figuras 25, 26 y 27, para solo los atributos “Nombres”, “Apellido Materno”, y “Apellido Paterno”, se puede desprender en la tabla 11 el resumen de los respectivos cálculos de indicadores que están bajo los requerimientos mínimos de exigencia.

Tabla	Indicadores por bajo los criterios
INDIVIDUO	Validez, consistencia, y exactitud
DOTACION	Consistencia
DATOS PERSONALES	Validez, consistencia y exactitud

Tabla 11 Resultados por tablas (Nombres y Apellidos)

Asimismo, en la tabla 12 se expone un resumen en cuanto a los indicadores de calidad que se encuentran por bajo los criterios definidos por el comité estratégico en cuanto a los atributos “Nombres”, “Apellido Materno”, y “Apellido Paterno” por identificador (RUT), desplegados en las figuras 28, 29 y 30, dentro de las 3 tablas a gobernar.

Tabla	Indicadores por bajo los criterios
INDIVIDUO	Validez, consistencia, y exactitud
DOTACION	Consistencia y exactitud
DATOS PERSONALES	Validez, consistencia y exactitud

*Tabla 12 Resultados por tablas (Nombres y Apellidos por identificador)*

Por otro lado, en la tabla 13 se despliega el resumen de los indicadores de calidad que se encuentran por bajo los criterios definidos por el comité estratégico en cuanto al atributo “Nombre” dentro de las 3 tablas a gobernar.

Tabla	Atributo	Indicadores por bajo los criterios
INDIVIDUO	NOMBRE	Validez, consistencia, y exactitud
DOTACION	NOMBRE	Consistencia, y exactitud
DATOS PERSONALES	NOMBRE	Validez, consistencia, y exactitud

*Tabla 13 Resultados atributo “Nombre”*

De igual modo, en la tabla 14 se despliega un resumen de los indicadores de calidad que se encuentran por bajo los criterios definidos por el comité estratégico en cuanto al atributo “Apellido Materno” dentro de las 3 tablas a gobernar.

Tabla	Atributo	Indicadores por bajo los criterios
INDIVIDUO	APELLIDO MATERNO	Validez y consistencia
DOTACION	APELLIDO MATERNO	Consistencia
DATOS PERSONALES	APELLIDO MATERNO	Validez, consistencia

*Tabla 14 Resultados atributo “Apellido Materno”*

Finalmente, en la tabla 15 se despliega un resumen de los indicadores de calidad que se encuentran por bajo los criterios definidos por el comité estratégico en cuanto al atributo “Apellido Paterno” dentro de las 3 tablas a gobernar.

Tabla	Atributo	Indicadores por bajo los criterios
INDIVIDUO	APELLIDO PATERNO	Validez, consistencia y exactitud
DOTACION	APELLIDO PATERNO	Consistencia
DATOS PERSONALES	APELLIDO PATERNO	Validez y consistencia

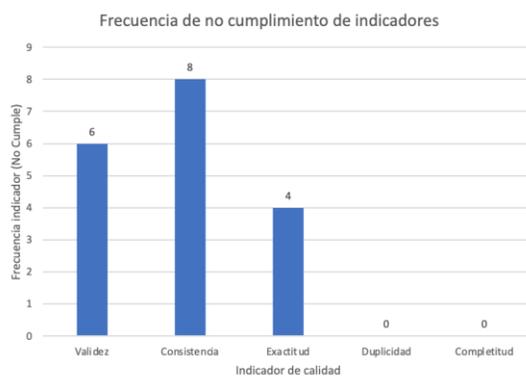
*Tabla 15 Resultados atributo “Apellido Paterno”*

## 8.5.4 Análisis resultados monitorización

En cuanto a los resultados de los indicadores obtenidos para los atributos relacionados en particular al promedio obtenido para nombres y apellidos desplegados en la tabla 11, se puede destacar que las tablas con mayor cantidad de indicadores bajo los requerimientos mínimos son “INDIVIDUO” y “DATOS PERSONALES”, donde específicamente, no hay cumplimiento de los criterios mínimos para los indicadores de validez, exactitud, y consistencia. Por otro lado, “DOTACION” solo posee un indicador fuera de los rangos establecidos, el que corresponde a consistencia. Asimismo, en cuanto a los resultados de los indicadores obtenidos dado el reporte por identificador (RUT) para los atributos relacionados en particular a nombres y apellidos desplegados en la tabla 12, se obtienen en general resultados inferiores para los indicadores de validez, consistencia y exactitud con respecto al análisis de promedios realizado anteriormente, y de esta manera, manteniendo estos indicadores por bajo los requerimiento mínimos exigidos por el comité estratégico, a excepción de la tabla “DOTACION” donde ahora existen 2 indicadores fuera de los criterios establecidos por el comité estratégico, siendo estos los de consistencia y exactitud.

Dados estos resultados, cabe destacar que esta disparidad de resultados entre “DOTACION” y las demás tablas se debe en estricto rigor al origen de los datos para las respectivas tablas, considerando aplicaciones que poseen distintos niveles de filtraje de validación a la entrada, además de la disparidad de procesos de limpieza de datos dentro de las tablas. En cuanto a “DOTACION”, los resultados obtenidos se deben a que es la tabla con menor cantidad de registros dentro de las tres tablas, y que parte de las aplicaciones que están conectadas para el ingreso de datos sobre esta tabla pasan por procesos de validación realizados manualmente, además de estar constantemente siendo monitoreada por áreas del negocio que tienen injerencia sobre la misma. Por otro lado, “INDIVIDUO” se presenta como la tabla más poblada de las tres a gobernar, donde existe una gran fluctuación de datos desde distintos orígenes, y por tanto, una gran cantidad de aplicaciones que accionan sobre esta tabla, lo que implica una difícil monitorización de entrada, considerando además que gran parte de las aplicaciones que están conectadas a esta tabla no poseen ningún tipo de validación de formato o comparación contra fuentes reales de información, afectando de esta manera indicadores como validez y exactitud. Finalmente, en cuanto a “DATOS PERSONALES”, se destaca que el ingreso de datos sobre esta tabla se hace desde muchas aplicaciones que no tienen validación de formato de los respectivos datos, afectando al indicador de validez, además de brindar gran libertad de inserción de información en los formularios a los usuarios, y sumando a inexistencia de procesos de comparación contra fuentes de información real, repercute en la existencia de nombres totalmente diferentes a los que cruzan con información correcta previamente validada, afectando de igual manera al indicador de exactitud. Cabe destacar que “DATOS PERSONALES” ha tenido procesos de limpiezas superficiales asociadas a formatos al igual que “INDIVIDUO”, pero considerando la disparidad entre las cantidades de atributos que poseen estas tablas, el impacto porcentual es más notorio sobre la tabla “DATOS PERSONALES”.

Dados los resultados expuestos en las tablas 13, 14, y 15, donde se despliegan los resultados obtenidos para cada atributo por separado, se puede categorizar los indicadores que tienen más frecuencia en el no cumplimiento de los criterios mínimos exigidos por el comité estratégico de gobierno de datos. Es así como en la tabla 16 se despliega las frecuencias de no cumplimiento para cada uno de estos indicadores sobre los atributos “Nombre”, “Apellido Materno”, y “Apellido Paterno”, dentro de las tablas “INDIVIDUO”, “DOTACION”, y “DATOS PERSONALES”



*Tabla 16 Frecuencia de no cumplimiento indicadores*

De la tabla 16 se puede desprender que el indicador con más frecuencia respecto al no cumplimiento de los requerimientos mínimos es la consistencia, seguido de validez y exactitud. Por otro lado, complementando en función de los resultados obtenidos por identificador (RUT), se puede establecer que los criterios más críticos transversalmente dentro de las tablas son los asociados a validez, consistencia y exactitud.

Los resultados obtenidos en cuanto al no cumplimiento para el indicador de consistencia se deben al origen de los datos para las respectivas tablas, donde las aplicaciones que ingresan datos, lo hacen con distintos niveles de validación para cada tabla. De igual manera, cabe destacar la disparidad entre los procesos de limpieza de los datos en las tablas, lo que tiene implicancia directa en el cruce de ellas. Finalmente, se destaca la inexistencia de procesos de integración entre estas correspondientes tablas, repercutiendo en que los problemas de consistencia se perpetúen a lo largo del tiempo.

Por otro lado, para el caso del indicador de validez, se puede recalcar que la alta frecuencia en el no cumplimiento de este indicador se debe a la falta de validación de datos transversalmente en la entrada hacia las bases de datos transaccionales dentro de gran parte de las aplicaciones que están actualmente conectadas a “INDIVIDUO” y “DATOS PERSONALES”, además de la inexistencia de limpieza y/o acciones correctivas masivas sobre estas.

Finalmente, en cuanto al indicador exactitud, también existe una relación directa con la entrada de datos sin validación, lo que hace que las fuentes reales de información crucen incorrectamente con los respectivos datos que se poseen actualmente dentro de las bases transaccionales de la Universidad de Chile, debido a la inexistencia de procesos de validación que verifiquen la exactitud de los datos.

### 8.5.5 Salida: Limpieza de datos

La segunda parte que contempla el proceso de despliegue dentro del plan de gobernanza implica una efectiva limpieza de datos en función de poder mejorar parte de los indicadores levantados para las tablas y atributos. De esta manera, para poder abordar esta problemática, se establecen tanto soluciones a corto plazo, como soluciones estratégicas que aborden la calidad del dato.

La primera solución integrada dentro de la proyección del plan en cuanto al corto plazo es establecer una limpieza de datos dentro de las tres tablas a gobernar, considerando los atributos en

cuestión. Esta limpieza está enfocada en poder abordar los indicadores más críticos levantados anteriormente, en específico, los indicadores de validez, consistencia y exactitud.

Por otro lado, la segunda solución integrada para abordar la calidad del dato está enfocada en poder generar una solución estratégica que pueda aplicarse a los nuevos ingresos de datos que se generen dentro de las bases transaccionales de la universidad, y así, luego de la limpieza histórica, poder garantizar que los datos que provengan desde aplicaciones sin ningún tipo de validación de formato, lleguen a las bases de datos con los formatos correctos, y que se consideren procesos de validación contra fuentes de información validas a cargo de algunas áreas involucradas.

### 8.5.5.1 Primera etapa de limpieza

Esta primera etapa tiene por objetivo generar una limpieza histórica dentro de las bases de datos transaccionales de la universidad en función de los indicadores de calidad críticos encontrados en la sección anterior dentro de las tablas y atributos a gobernar. Es así como los indicadores a abordar serán los de validez, consistencia y exactitud. Cabe destacar que los respectivos análisis para las limpiezas serán en función de los resultados obtenidos por identificador (RUT).

#### 8.5.5.1.1 Primera etapa de limpieza: Validez

En cuanto al indicador de validez, la solución propuesta es poder otorgar el formato correcto a los atributos dentro de las tablas a gobernar, a nivel histórico. Para esto, el formato definido para la respectiva limpieza son los acordados en la etapa de perfilamiento, y por tanto, que los atributos correspondiente a nombres y apellidos mantengan la primera letra en mayúscula, y las demás en minúsculas, tal como se expresa en la tabla 17.

FORMATO_CORRECTO	FORMATO_INCORRECTO
1 Marco Andrés Olguín Decencier	MARCO ANDRES OLGUIN DECENCIER

Tabla 17 Estandarización nombres y apellidos

En primer lugar, es necesario poder levantar los respectivos resultados obtenidos para el indicador de validez en el *dashboard* de calidad, considerando las tablas y atributos a gobernar en cuanto al análisis por identificador. En la figura 28, el indicador de validez para la tabla “INDIVIDUO” alcanza el 37%. Por otro lado, para la figura 29 el indicador de validez para la tabla “DOTACION” alcanza el 86%. Finalmente, para la figura 30 el indicador de validez para la tabla “DATOS PERSONALES” alcanza el 66%.

Dados los resultados desplegados, se puede ver que existen márgenes de mejora de 63 puntos porcentuales para “INDIVIDUO”, 34 puntos porcentuales para “DATOS PERSONALES”, y 14 puntos para “DOTACION”. Luego de levantar los respectivos cálculos para la validez dentro de las tablas y atributos a gobernar, se procede a generar la limpieza sobre estas, utilizando una función SQL desarrollada para mantener los formatos establecidos para los atributos en cuestión, expresada en el anexo L, la que permite:

1. Devolver la cadena de valor empezando la primera letra con mayúsculas y el resto en minúsculas, mediante la función que provee ORACLE llamada “INITCAP”.

2. Eliminar espacios antes, entre, y después del nombre y/o apellido, mediante la función que provee ORACLE llamada “TRIM”.
3. Mediante el condicional “IF”, y la función que ofrece ORACLE “REPLACE”, se identifican todos los nombres y/o apellidos compuestos, y así permitir excepciones para el caso de la función “INITCAP”.

De esta manera, al definir la función para mantener los formatos dentro de las tablas a gobernar, se procede a accionar la limpieza mediante la función que provee ORACLE “UPDATE” dentro de las tablas a gobernar. A modo referencial, se despliega el antes y después de los registros para la tabla “INDIVIDUO” en la tabla 18, donde en las 3 primeras columnas se expresan los atributos antes de la limpieza, y luego, en las 3 últimas columnas se despliegan los nombres y apellidos luego de la respectiva limpieza, reflejando así la normalización en función del estándar definido.

	NOMBRE_UCHILE	PATERNO_UCHILE	MATERNO_UCHILE	NOMBRES_REGISTROC	PATERNO_REGISTROC	MATERNO_REGISTROC
1	MACARENA ALEJANDRA	FÉREZ	SALGADO	Macarena Alejandra Pérez		Salgado
2	SALVADOR ALEJANDRO	FERNÁNDEZ	LAGOS	Salvador Alejandro Fernández		Lagos
3	OLIVIA CRISTINA	MUÑOZ	NÚÑEZ	Olivia Cristina Muñoz		Núñez
4	FABIOLA DEL PILAR	ALVEAL	BERROCAL	Fabiola del Pilar Alveal		Berrocal
5	WILLIAMS ALFREDO	JARA	GUAJARDO	Williams Alfredo Jara		Guajardo
6	MARÍA CECILIA	CRISOSTO	ELGUETA	María Cecilia Crisosto		Elgueta

*Tabla 18 Estandarización nombres y apellidos luego de limpieza (INDIVIDUO)*

Cabe destacar que la limpieza en función de formatos realizada dentro de las tablas a gobernar contempla todos los atributos que efectivamente no cumplen con los formatos establecidos para nombres y apellidos dentro de las respectivas tablas a gobernar, sin considerar los campos que tengan registros nulos.

Finalmente, luego de generar la respectiva limpieza sobre las tablas, se logra poder formatear los registros en función de los estándares acordados en la etapa de perfilamiento, y así lograr aumentar el indicador de validez en 63 puntos porcentuales para la tabla “INDIVIDUO”, 14 para “DOTACION”, y finalmente 34 para “DATOS PERSONALES”, obteniendo así mejoras considerables con respecto a lo levantado al inicio de esta memoria. En la tabla 19 se despliega el crecimiento del indicador de validez de las tablas a gobernar. Por otro lado, en los anexos M, N, y O se encuentran las gráficas obtenidas desde el panel de calidad desarrollado en PowerBI.

Tabla	Validez pre limpieza	Validez post limpieza	Puntos porcentuales de mejora (Delta)
INDIVIDUO	37%	100%	63
DOTACION	86%	100%	14
DATOS PERSONALES	66%	100%	34

*Tabla 19 Porcentaje de mejora validez post limpieza*

De esta manera, se alcanza para este indicador porcentajes de cumplimiento de 100% para todas las tablas abordadas, logrando así brindar el formato correcto a los respectivos datos históricos alojados en estas. Finalmente, dadas las respectivas acciones reparatorias, se cumple con la primera etapa de limpieza, abordando los datos históricos para el indicador de validez dentro de las tablas

a gobernar, logrando de esta manera estandarizar la totalidad de los datos alojados a los formatos establecidos en la etapa de perfilamiento.

#### 8.5.5.1.2 Primera etapa de limpieza: Exactitud

En cuanto al indicador de exactitud, la solución propuesta para abordar la limpieza de éste es poder generar algún acceso a una fuente real de información, previamente validada, que permita poder cruzar registros contra las bases transaccionales de la universidad, para luego efectuar las eventuales medidas reparatorias. De esta manera, para generar la respectiva limpieza se utiliza la tabla STAGE referenciada en el proceso de carga de datos del Registro Civil, la que contiene toda la información que emite el Registro Civil, y que cruza por identificador (RUT) con respecto a las tablas a gobernar, y de esta manera, luego de efectivamente cargar los datos obtenidos desde el registro civil dentro de la tabla STAGE, se generan los respectivos análisis pre-limpieza en cuanto a los atributos a gobernar.

Asimismo, es necesario poder visualizar los respectivos resultados levantados para el indicador de exactitud por el *dashboard* de calidad en las tablas y atributos a gobernar dado al análisis por identificador (RUT). En primer lugar, en cuanto a la figura 28, el indicador de exactitud para la tabla “INDIVIDUO” alcanza el 45%. Por otro lado, en la figura 29 el indicador de exactitud para la tabla “DOTACION” alcanza el 58%. Finalmente, en la figura 30 el indicador de exactitud para la tabla “DATOS PERSONALES” alcanza el 37%. De esta manera, el indicador de exactitud para estas tablas expone resultados bajo los criterios mínimos exigidos por el comité estratégico de gobierno de datos.

Desglosando el indicador de exactitud, que hace referencia al cruce con respecto a información del registro civil, cabe destacar que existen 2 distinciones: cruce correcto y cruce incorrecto. Para la primera, hace referencia a los registros que cruzan contra el registro civil, es decir, que el identificador (RUT) se encuentre tanto en las bases de la universidad como en el registro civil, y que los registros asociados a nombres y apellidos se encuentren iguales. Por otro lado, el cruce incorrecto hace referencia a registros que cruzan contra el registro civil, es decir, que el identificador (RUT) se encuentre tanto en las bases de la universidad como en el registro civil, pero que alguno de los atributos a gobernar sea diferente. Dentro de los atributos que cruzan de manera incorrecta, existen 3 tipos, los que corresponden a diferencias por tildes y/o signos, nombres y/o apellidos diferentes, y atributos que poseen *Null* dentro de las bases transaccionales de la universidad.

El primero (diferencias por tildes) corresponde a atributos asociados a nombres y/o apellidos que solo se diferencian contra el registro civil por la existencia de tildes. Este caso se ejemplifica en la tabla 20, se puede apreciar que dentro de las bases de la universidad el nombre asociado corresponde a “Marco Andrés Olguín Decencier”, mientras que en el registro civil se almacena como “Marco Andres Olguin Decencier”, siendo el tilde la única diferencia entre estos registros, y por tanto, siendo candidato a limpieza. Cabe destacar que estas diferencias por tildes para la eventual limpieza pueden ocurrir tanto para nombres como apellidos.

BASES_UCHILE	REGISTRO_CIVIL
1 Marco Andrés Olguín Decencier	Marco Andres Olguin Decencier

Tabla 20 Ejemplificación limpieza por tildes

El segundo caso (nombres y/o apellidos diferentes) corresponde a atributos donde la diferencia contra el registro civil no sea exclusivamente por tildes, es decir, que considere atributos que se diferencian ya sea por alguna letra diferente, o diferencias totales en los nombres, tal como se aprecia en la tabla 21, donde las tres primeras columnas corresponden a los registros de nombre, apellido paterno, y apellido materno que pertenecen a las bases de datos de la Universidad de Chile, y por otro lado, las últimas 3 columnas corresponden a los registros de nombre, apellido paterno, y apellido materno dictado por el registro civil, donde a modo ejemplo, tomando el primer registro, se puede apreciar que dentro de las bases de la universidad el nombre asociado corresponde a “Alejandro”, mientras que en el registro civil se almacena como “José León”, diferenciándose completamente. Siguiendo el ejemplo, en el segundo atributo de la tabla, dentro de las bases de la universidad el apellido paterno asociado corresponde a “Ramos”, mientras que en el registro civil se almacena como “Ciapa”, existiendo una diferencia total en el respectivo apellido.

NOMBRE_UCHILE	PATERNO_UCHILE	MATERNO_UCHILE	NOMBRES_REGISTROC	PATERNO_REGISTROC	MATERNO_REGISTROC
Alejandro	Munoz	Nunez	José León	Elgueta	Catalán
Manuel Arturo	Ramos	Berrocal	Guillermo Danilo	Ciapa	Pérez

Tabla 21 Ejemplificación nombres y/o apellidos diferentes

Finalmente, el tercer caso (atributos que poseen *Null*) corresponde a atributos que en las bases transaccionales de la universidad posea *Null* en algunos de los atributos a gobernar, y en el registro civil se obtenga un campo diferente de *Null*.

Es así como las medidas reparatorias dentro de los alcances de esta memoria en función de los atributos cruzados con el registro civil tienen por objetivo generar una limpieza por tildes y/o signos, es decir, poder identificar y reparar los registros que actualmente cruzan incorrectamente por esta razón. Los demás casos de cruce incorrecto involucran la responsabilidad de una gran cantidad de áreas que tienen custodia sobre los datos y, por tanto, no siendo factible la realización de estas limpiezas dentro del alcance de la presente memoria.

De esta manera, a modo de análisis, se generan las respectivas *queries* de cálculo, para así finalmente visualizar la cantidad de registros que cruzan incorrectamente por esta razón (Diferencia en tildes), y que eventualmente podrían ser candidatos a limpieza. En la tabla 22 se despliegan los respectivos resultados obtenidos para los cálculos de la cantidad de registros que cruzan con respecto al registro civil, y los que eventualmente serían reparados con esta limpieza.

Tabla	Registro que cruzan (Correctamente e incorrectamente)	Registros que se repararán tras limpieza de tildes
INDIVIDUO	3.656.675	1.462.670
DOTACION	112.595	36.030
DATOS PERSONALES	309.446	71.172

Tabla 22 Tabla de resultados por tildes

Tal como expone la tabla 22, y llevando a porcentajes los cálculos realizados, se puede destacar que en la tabla “INDIVIDUO” existe un margen de mejora de 40 puntos porcentuales con respecto a atributos que poseen en algunos de los campos diferencias por tildes con respecto al registro civil. Por otro lado, se expone que en la tabla “DOTACION” existe un margen de mejora de 32 puntos

porcentuales, y finalmente en la tabla “DATOS PERSONALES” existe un margen de mejora de 23 puntos porcentuales.

Cabe destacar que los registros que cruzan con respecto al registro civil para cada una de las tablas a gobernar no corresponde a la totalidad de registros que contienen estas tablas, dada la existencia de registros que efectivamente no cruzan con el registro civil, y que debido a esto, preliminarmente no se consideran candidatos a ninguna de las tres limpiezas mencionadas para abordar el indicador de exactitud, situación que se refleja en que, a modo referencial, actualmente existen aproximadamente 4,8 millones de registros para la tabla “INDIVIDUO”, pero los que actualmente cruzan, ya sea correcta o incorrectamente, son 3,6 millones, situación que es debida al ingreso de RUTS incorrectos y/o inválidos dentro de estas respectivas tablas, tal como se destaca en la tabla 23, abarcando el 15% para “INDIVIDUO”, el 10% para “DOTACION”, y el 23% para “DATOS PERSONALES”.

ID	NOMBRES	PATERNO	MATERNO
357 0 86935414		Rut Inexistente	(null)
358 0 89674500		Sotrul Limitada	(null)
359 0 91468000		Rut Inexistente	(null)
360 0 95892254		Rut Inexistente	(null)
361 0 96512070		Rut Inexistente	(null)
362 0#¿NOMBRE? C.		Ocampo	(null)

*Tabla 23 Registros que no cruzan con el RC*

Es así como se inicia el proceso de limpieza del indicador de exactitud dentro de las tablas y atributos a gobernar, donde inicialmente se crean tres tablas temporales en función de las tablas a gobernar, donde en estas se cargan todos los registros que cruzan contra el registro civil (Cruce correcto e incorrecto) para los atributos asociados a nombres y apellidos, cargando en sus respectivas columnas los datos alojados en las tablas a gobernar, y por otro lado, los obtenidos desde el registro civil. A modo referencia, en la tabla 24 se expone el caso para la tabla “INDIVIDUO”, donde para esta nueva tabla temporal creada, se establece que en las primeras tres columnas se cargan los nombres y apellidos que se alojan en “INDIVIDUO”, y en las tres últimas columnas se cargan los registros obtenidos desde el registro civil. Cabe destacar que, dentro de las tres últimas columnas con los registros obtenidos desde el registro civil, solo se cargan efectivamente los atributos que difieren por tildes, dejando los que cruzan correctamente como *Null*, y de esta manera, identificando solo los registros que cruzan incorrectamente, y en qué respectivo atributo.

IND_NOMBRES	IND_PATERNO	IND_MATERNO	IND_NOMBRES_FIX	IND_PATERNO_FIX	IND_MATERNO_FIX
22 Claudio Cesar	Briceño	Perez	Claudio César	(null)	Pérez
23 Alejandra Paz	Samaniego	Acuna	(null)	(null)	Acuña

*Tabla 24 Tabla temporal de cruce (Exactitud)*

Luego de poblar satisfactoriamente estas tablas temporales, se utiliza la función que ofrece ORACLE “MERGE” para finalmente migrar los datos que efectivamente se encontraban con diferencias exclusivamente por tildes hacia las tablas “INDIVIDUO”, “DOTACION”, y “DATOS PERSONALES”. A modo referencia, en el anexo P se encuentra el respectivo procedimiento almacenado para realizar la migración de la tabla “INDIVIDUO”.

Finalmente, luego de generar la respectiva limpieza de tildes sobre las tablas, se logra aumentar los indicadores de exactitud en 40 puntos porcentuales para la tabla “INDIVIDUO”, 32 para “DOTACION”, y finalmente 23 para “DATOS PERSONALES”, obteniendo así mejoras

considerables con respecto a lo levantado al inicio de esta memoria. En la tabla 24 se despliega el crecimiento del indicador de exactitud de las tablas a gobernar. Por otro lado, en los anexos Q, R y S se encuentran las gráficas obtenidas desde el panel de calidad desarrollado en PowerBI.

Es así como se puede obtener mejoras considerables dentro de “INDIVIDUO” y “DOTACION”, siendo la diferencia por tildes con respecto al Registro Civil la razón principal de cruce insatisfactorio para los registros asociados a nombres y apellidos, logrando para estas tablas alcanzar los respectivos criterios mínimos exigidos por el comité estratégico de datos. Por otro lado, con respecto a “DATOS PERSONALES”, si bien se logran mejoras considerables para esta tabla, no fueron suficientes para alcanzar el criterio mínimo de cumplimiento, y de esta manera, se deduce que gran parte de los registros para esta tabla cruzan incorrectamente por razones asociadas a nombres y/o apellidos diferentes, o existencias de *Nulls* en esta tabla, donde la limpieza de estos casos quedará propuesta como trabajo futuro.

Dada la tabla 25, en cuanto a los atributos que siguen sin dar cumplimiento al indicador de exactitud luego de la limpieza por tildes, por un lado, para “INDIVIDUO” se puede destacar que el 15% restante, es decir, exactamente 548.501 registros, cruzan incorrectamente debido a razones asociadas a nombres y/o apellidos diferentes, o existencia de *Nulls* para registros dentro de la respectiva tabla. Para “DOTACION”, el 10% restante, correspondiente a 11.259 registros, cruzan incorrectamente debido a razones asociadas a nombres y/o apellidos diferentes, o existencia de *Nulls* para registros dentro de la respectiva tabla. Finalmente, para “DATOS PERSONALES”, el 40% restante, correspondiente a 123.778 registros, cruzan incorrectamente debido a razones asociadas a nombres y/o apellidos diferentes, o existencia de *Nulls* para registros dentro de la respectiva tabla. Cabe destacar que la principal razón para que el indicador de exactitud dentro de “DATOS PERSONALES” no subiera tanto comparado a las demás tablas, es debido a que muchas de las aplicaciones que ingresan data a esta tabla no poseen ningún tipo de validación, además de brindar gran libertad de inserción de información a los respectivos usuarios, implicando en que exista una gran cantidad de campos en *Null*, o con nombres totalmente diferentes a los que indica el Registro Civil.

Tabla	Exactitud pre limpieza	Exactitud post limpieza	Puntos porcentuales de mejora (Delta)
INDIVIDUO	45%	85%	40
DOTACION	58%	90%	32
DATOS PERSONALES	37%	60%	23

Tabla 25 Porcentaje de mejora exactitud post limpieza

### 8.5.5.1.3 Etapa de limpieza indicador consistencia

Tal como se levantó en el *dashboard* de calidad, uno de los problemas más transversales en cuanto a indicadores de calidad era el asociado a consistencia, y de esta manera, en función de dar cumplimiento a este indicador, las respectivas limpiezas de datos históricos para los indicadores de validez y exactitud dentro las tablas seleccionadas se realizaron manteniendo la misma estructura, es decir, la función de formateo aplicada para mantener formatos dentro de las tablas (Validez), y el respectivo proceso de limpieza por tildes (Exactitud) se aplica de la misma manera transversalmente dentro de las tablas, logrando finalmente estandarizar de manera unificada los

datos dentro de cada una de la tablas, en específico, “INDIVIUO”, “DOTACION”, Y “DATOS PERSONALES”.

Asimismo, luego de poder haber generado las respectivas limpiezas históricas enfocadas en abordar los indicadores de exactitud y validez, y levantar las respectivas mejoras para cada uno de estos, siguiendo la hipótesis anteriormente planteada en cuanto a la estandarización de datos, finalmente se obtienen mejoras con respecto al indicador de consistencia para las 3 tablas a gobernar. De esta manera, se logra aumentar los indicadores de consistencia en 23 puntos porcentuales para la tabla “INDIVIDUO”, 21 puntos para “DOTACION”, y finalmente 20 puntos porcentuales para “DATOS PERSONALES”, obteniendo así mejoras considerables con respecto a lo levantado al inicio de esta memoria. En la tabla 26 se despliega el crecimiento del indicador de consistencia de las tablas a gobernar. Por otro lado, en los anexos T, U y V se encuentran las gráficas obtenidas desde el panel de calidad desarrollado en PowerBI.

Tabla	Consistencia pre limpieza	Consistencia post limpieza	Puntos porcentuales de mejora (Delta)
INDIVIDUO	55%	78%	23
DOTACION	64%	85%	21
DATOS PERSONALES	51%	71%	20

*Tabla 26 Porcentaje de mejora consistencia post limpieza*

Dada la respectiva limpieza, se puede obtener mejoras considerables dentro de “INDIVIDUO” y “DOTACION”, y “DATOS PERSONALES”, sin embargo, aún existen márgenes de mejora para este indicador. A modo análisis, en la tabla 27 se despliega las respectivas mejoras en puntos porcentuales de cruce (Consistencia) entre las respectivas tablas luego de las limpiezas. De esta manera, se destaca que luego de las limpiezas, “INDIVIDUO” mejoró en 18 puntos porcentuales de consistencia para el cruce con “DOTACION”, y 17 para “DATOS PERSONALES”. Por otro lado, “DOTACION” mejoró 18 puntos porcentuales de consistencia para el cruce con “INDIVIDUO”, y 19 para “DATOS PERSONALES”. Finalmente, “DATOS PERSONALES” refleja mejoras de 17 puntos porcentuales de consistencia para el cruce con “INDIVIDUO”, y 19 para “DOTACION”.

Tabla	INDIVIDUO	DOTACION	DATOS PERSONALES
INDIVIDUO	-	18	17
DOTACION	18	-	19
DATOS PERSONALES	17	19	-

*Tabla 27 Mejoras puntos porcentuales consistencia entre tablas*

De esta manera, se destaca que las respectivas limpiezas realizadas para abordar los indicadores de validez y exactitud tuvieron efectos transversales y uniformes dentro de las respectivas tablas para el indicador de consistencia, donde estas mejoras se mantuvieron dentro del espectro de 17-19 puntos porcentuales. Es así como se desprende que los problemas asociados a formatos y tildes eran un porcentaje importante de cruce incorrecto para antes de la limpieza, y que luego de esta, mejoró transversalmente los porcentajes de cruce entre estas tablas. Cabe destacar que los registros que luego de la limpieza no cruzan correctamente para el indicador de consistencia, y en particular,

22% para “INDIVIDUO”, 15% para “DOTACION”, y 29% para “DATOS PERSONALES”, tal como detalla la tabla 26, corresponden a registros que poseen diferencias asociadas a nombres y/o apellidos diferentes entre tablas, omisión de segundos nombres, *nulls*, entre otros, los que corresponden a casos que deben ser tratados directamente por las áreas que poseen custodia sobre estos datos.

#### **8.5.5.1.4 Segunda etapa de limpieza: Validez**

La segunda etapa de limpieza para el indicador de validez hace referencia a la acción estratégica que se propone para poder abordar las problemáticas actuales con respecto a la entrada de datos sin cumplimiento de formato dentro de las tablas transaccionales de la Universidad de Chile.

Uno de los problemas estratégicos levantados dentro de la justificación del problema para la presente memoria es el ingreso de datos sin formato dentro de las bases transaccionales de la universidad, teniendo como origen las distintas aplicaciones que actualmente se encuentran activas para el ingreso de datos, las que en su mayoría permiten ingreso de información sin ningún tipo de validación en los respectivos formularios que cada una posee. De manera paralela a la limpieza de los datos históricos (Primera etapa de limpieza) sobre las tablas, se establece un proceso que permita poder dar formato a los nuevos datos entrantes luego de la primera limpieza, provenientes desde aplicaciones que no poseen validación de formato, para así no permitir ingreso de datos sin validación de formato, estableciendo así una estrategia transversal de aseguramiento de formato.

El proceso que se plantea para abordar de manera estratégica la entrada de datos a las tablas a gobernar es la implementación de un Trigger SQL, que se explicita a modo referencial en el anexo W, el que permite ingresar como regla dentro de las respectivas tablas a gobernar un gatillador de formato a la entrada de datos desde diversas aplicaciones, es decir, cada vez que ingresen datos nuevos dentro de las tablas a gobernar, se genere un gatillador de formato, que permita ingresar el formato correcto a los datos entrantes dentro de las tablas a gobernar, utilizando la función descrita en el anexo L, previamente utilizada para la respectiva limpieza histórica. El proceso que describe el gatillador de formato se diagrama en el anexo X. Este proceso parte con el ingreso de información de los usuarios de aplicaciones sobre los respectivos formularios que proveen para el ingreso de información sobre la misma, para así inmediatamente ingresar estos datos dentro de las bases transaccionales de la universidad. A la llegada de la *data* dentro de las tablas, se gatilla el procedimiento Trigger ingresado como regla, el que mediante la función de estandarización definida en el anexo L, permite formatear los atributos asociados a nombres y apellidos, estableciendo el estándar definido en la etapa de perfilamiento. Asimismo, estos datos ingresan a las respectivas tablas respetando los formatos establecidos, para finalmente terminar el proceso.

Es así como se logra poder establecer una estrategia de entrada para el formateo de datos, y así poder trabajar transversalmente el indicador de validez, manteniendo en función del tiempo las mejoras realizadas en el proceso de limpieza histórica. Finalmente, se aborda de esta manera tanto táctica como estratégicamente la validez dentro de los atributos y tablas a gobernar.

#### **8.5.5.1.5 Segunda etapa de limpieza: Exactitud**

La segunda etapa de limpieza para el indicador de exactitud hace referencia a la acción estratégica que se propone para poder abordar las problemáticas actuales con respecto a la entrada de datos sin validación desde una fuente real de información dentro de las tablas a gobernar. Para esta segunda

etapa de limpieza, se propone un complemento a la segunda etapa de limpieza asociada al indicador de validez, donde además de poder evaluar el formato de los datos entrantes a las bases de datos, se pueda generar una respectiva validación contra el registro civil, y así abordar estratégicamente el indicador de exactitud. Esta validación propuesta como solución permite identificar nuevos datos entrantes dentro de las tablas a gobernar, y poder compararlos con los respectivos datos que emite oficialmente el registro civil para el identificador a consultar, y así generar una identificación oportuna de los registros que eventualmente no crucen de manera correcta para los atributos a gobernar.

El proceso de validación parte con la implementación de un Trigger SQL, el que a modo referencial se expresa en el anexo Y, el que permita ingresar como regla dentro de las respectivas tablas a gobernar que los nuevos datos entrantes sean enviados a una tabla temporal, expresada en la tabla 28, la que tiene por función principal poder almacenar estos respectivos datos entrantes a las tablas a gobernar. Por otro lado, se implementa un procedimiento almacenado que consulte esta respectiva tabla temporal con los registros cargados, y genere un llamado a la API del registro civil, cargando los datos que tiene el registro civil para los registros consultados, y así finalmente genere un cruce de información. En otras palabras, se envían los datos entrantes a las tablas a gobernar hacia una tabla temporal que da “seguimiento” a estos, para luego compararlos contra el registro civil.

En el anexo Z se explicita el respectivo BPMN que detalla el proceso de validación de *data* entrante a las tablas a gobernar. El proceso parte con el ingreso de información de los usuarios de aplicaciones sobre los respectivos formularios que proveen para el ingreso de información sobre la misma, e inmediatamente ingresar esta *data* dentro de las bases transaccionales de la universidad. A la llegada de la data dentro de las tablas, se gatilla el procedimiento Trigger, el que ingresa estos datos entrantes a una tabla temporal. Esta tabla temporal tiene por objetivo poder identificar cualquier tipo de acción sobre las tablas a gobernar, distinguiendo como columnas a los nombres y apellidos antes y después del cambio y/o inserción, en qué tabla fue modificado, el tipo de acción sobre la tabla, fecha de modificación y/o inserción, el respectivo usuario con cuenta en bases de datos y/o aplicación que ha realizado el cambio, dato extraído desde el Registro Civil, y estado de actualización, trazando tanto en tiempo como origen los datos a la entrada de las tablas a gobernar. Cabe destacar que los registros son cargados desde el Registro civil mediante el script descrito en el anexo G. Luego, existen 2 casos posibles; en el caso de que el dato entrante cruce correctamente con el registro civil, termina el proceso, y por otro lado, en caso de que el registro cruce incorrectamente con la información emitida por el Registro Civil, se agrega una marca sobre el respectivo atributo dentro de la tabla, en específico la columna “ESTADO\_REVISION” (ESTADO\_REVISION posee valor de Ok en el caso que cruce correctamente, y NOOK en el caso que no cruce correctamente), donde finalmente se notifica a las áreas involucradas para la efectiva reparación de este atributo. A modo ejemplo, en la tabla 28 se encuentra ejemplificada la tabla temporal para el caso del atributo “NOMBRE”.

❖ RUT	❖ NOMBRES	❖ ORIGEN_DATO	❖ NOMBRES_ANT	❖ OPERACION	❖ USUARIO	❖ NOMBRES_SRCEI	❖ ESTADO_REVISION
1 0113394277	Sofia Eliana	INDIVIDUO	SOFIA ELIANA	UPDATE	CLAUDIABRAVO	(null)	(null)
2 0236560651	Ignacia Catalina	INDIVIDUO	Ignacia	UPDATE	FABIOLAGUERRA	(null)	(null)
3 0259525322	Nestor Manuel	INDIVIDUO	Nestor	UPDATE	FABIOLAGUERRA	(null)	(null)
4 0153605645	Felipe Andrés	INDIVIDUO	FELIPE ANDRES	UPDATE	RGUZMANBOZO_TYG	(null)	(null)
5 0179586681	Ignacio Tomás	INDIVIDUO	Ignacio Tomas	UPDATE	RGUZMANBOZO_TYG	(null)	(null)
6 0194167997	César Mauricio	INDIVIDUO	CESAR MAURICIO	UPDATE	RGUZMANBOZO_TYG	(null)	(null)
7 0154141294	Rodrigo Antonio	INDIVIDUO	Rodrigo	UPDATE	RGUZMANBOZO_TYG	(null)	(null)
8 0176010649	Florencia Julia	INDIVIDUO	Florencia Julia	UPDATE	SUG	(null)	(null)
9 0153372756	Bárbara Alejandra	DOTACION	Bárbara Alejandra	UPDATE	ERTIRADO	(null)	(null)

Tabla 28 Tabla de paso (Exactitud)

De esta manera, se logra poder establecer una estrategia de entrada para abordar el indicador de exactitud, y por tanto la respectiva validación de nueva *data* entrante dentro de las tablas a gobernar, para así, dada la limpieza de datos históricos realizada en función de abordar el indicador de exactitud, establecer una estrategia mixta de validación de *data* contra el Registro Civil.

Finalmente, dadas las acciones realizadas tanto de monitorización de los atributos y tablas a gobernar como la limpieza táctica y estratégica en función de los indicadores críticos levantados por el panel de calidad, se concluye la realización del último paso que comprende la implementación de la gobernanza dentro de la Universidad de Chile, dando cumplimiento al último objetivo específico pactado para esta memoria.

## 9 CONCLUSIONES

En el capítulo 6 de esta memoria se enuncian los respectivos objetivos que proyecta esta, destacando 5 objetivos específicos que dan cumplimiento al objetivo general pactado. Estos 5 objetivos específicos fueron trabajados y alcanzados para efectos de esta memoria, y en la presente sección se detallarán brevemente los resultados para cada uno de estos.

1. Se modeló y estructuró un plan de trabajo para la implementación de la gobernanza de datos sobre la Universidad de Chile (Metodología), estructurando además los respectivos comités a nivel organizativo, los que permitan generar toma de decisiones, para posteriormente ser aprobadas por los máximas autoridades de la universidad, y de esta manera, estableciendo los respectivos lineamientos organizativos para una correcta implementación del plan de gobernanza, generando las bases tomadoras de decisiones para la acción tomada dentro de la presente memoria, y finalmente dando cumplimiento del 100% para el objetivo en cuestión.
2. Dentro del plan desarrollado para la implementación de la gobernanza de datos en la Universidad de Chile se identifican los respectivos datos a gobernar para efectos de los alcances de esta memoria, en específico los atributos relacionados a nombres y apellidos dentro de las tablas de datos que más transaccionan estos. Esta identificación y posterior priorización para su elección se basó principalmente en la importancia que refiere los nombres y apellidos dentro de gran cantidad de procesos que actualmente posee en ejecución la universidad, además del volumen de uso de este dato dentro de estos mismos. Por otro lado, se define el alcance de estos atributos a las tablas de datos “INDIVIDUO”, “DOTACION”, y “DATOS PERSONALES”, las que destacan por ser las tablas que más transaccionan estos respectivos datos. De esta manera se da cumplimiento del 100% para el objetivo en cuestión.
3. Por otro lado, si bien la presente memoria tiene un especial énfasis en calidad del dato, no deja de lado otras aristas importantes que convergen la gobernanza de datos. Es así como dentro de lo proyectado y finalmente realizado dentro de los alcances, se destaca un

levantamiento sobre la actual arquitectura del dato sobre aplicaciones, reflejándose esto en los diagramas realizados para las distintas aplicaciones que posee la Universidad de Chile, las que transaccionan datos sobre las bases de datos institucionales, además de obtener una matriz CRUD en forma de reporte, la que permite visualizar de manera general y particular los respectivos permisos que poseen las cuentas creadas para aplicaciones y bases de datos, que finalmente influyen directamente sobre el actual flujo de datos. De esta manera, se da cumplimiento del 100% para el objetivo en cuestión.

4. El perfilado de datos es una arista fundamental sobre la correcta gobernanza, dado que es la acción de poder levantar toda la información necesaria para generar un correcto levantamiento y análisis de los problemas. Es así como dentro de lo realizado en este plan de gobernanza, se destaca la estandarización con respecto a formatos de los datos a gobernar, estableciendo un formato transversal dentro de la institución. Por otro lado, y en función de poder levantar los actuales problemas asociados a calidad del dato, se establecen indicadores para medir la actual calidad de dato bajo 5 criterios de estudio definidos por DAMA; Exactitud, consistencia, validez, unicidad y completitud. Es así como se cumple en totalidad el perfilamiento de datos, levantando formatos e indicadores de calidad, y de esta manera, se da cumplimiento del 100% para el objetivo en cuestión.
5. Finalmente, se levantan tanto soluciones de monitoreo como de limpieza para poder abordar las actuales problemáticas de datos que se reflejaron en los indicadores definidos para su respectivo uso. Es así como se desarrolló un completo panel de calidad elaborado en PowerBI, el que permitió identificar los problemas asociados a la calidad del dato sobre los atributos asociados a nombres y apellidos dentro de las tablas a gobernar. Los resultados obtenidos para estos atributos fueron los siguientes (Promedios de los resultados obtenidos entre los 3 atributos gobernados):
  - Para la tabla “INDIVIDUO”: Completitud 89%, Validez 50%, Consistencia 71%, Unicidad 100%, y Exactitud 72%
  - Para la tabla “DOTACION”: Completitud 100%, Validez 100%, Consistencia 73%, Unicidad 100%, y Exactitud 86%
  - Para la tabla “DATOS PERSONALES”: Completitud 94%, Validez 73%, Consistencia 58%, Unicidad 100%, y Exactitud 72%

Por otro lado, en términos de análisis por identificador (RUT), se despliegan los siguientes resultados asociados a las tablas a gobernar.

- Para la tabla “INDIVIDUO”: Completitud 84%, Validez 46%, Consistencia 61%, Unicidad 100%, y Exactitud 48%
- Para la tabla “DOTACION”: Completitud 100%, Validez 97%, Consistencia 68%, Unicidad 100%, y Exactitud 60%

- Para la tabla “DATOS PERSONALES”: Completitud 83%, Validez 69%, Consistencia 55%, Unicidad 100%, y Exactitud 38%

Asimismo, en función del levantamiento de indicadores, se definieron e implementaron estrategias de limpieza para los indicadores de calidad más críticos dentro de las tablas que más transaccionan los datos a gobernar, en particular a los indicadores de validez, exactitud y consistencia, logrando los siguientes resultados dentro del análisis por identificador (RUT).

- Mejoras en puntos porcentuales de indicadores dentro de la tabla “INDIVIDUO”: Validez 63 puntos porcentuales, Consistencia 23 puntos porcentuales, y Exactitud 40 puntos porcentuales.
- Mejoras en puntos porcentuales de indicadores dentro de la tabla “DOTACION”: Validez 14 puntos porcentuales, Consistencia 21 puntos porcentuales, y Exactitud 32 puntos porcentuales.
- Mejoras en puntos porcentuales de indicadores dentro de la tabla “DATOS PERSONALES”: Validez 34 puntos porcentuales, Consistencia 20 puntos porcentuales, y Exactitud 23 puntos porcentuales.

Finalmente se implementan estrategias de formateo y validación a la entrada de los datos hacia las tablas a gobernar, abordando los nuevos datos que eventualmente entrarán a las bases de datos de la universidad, estableciendo así una estrategia mixta que permita tanto reparar los datos históricos, como asegurarse de que la nueva data entrante pase por procesos de formateo y validación.

Es así como se logra poder cumplir los 5 objetivos específicos pactados al inicio de esta memoria, y que convergen en el cumplimiento del objetivo general acordado dentro de los alcances esperados. De esta manera, y en base al modelamiento e implementación del plan en la Universidad de Chile, se establece lineamientos y bases sólidas para la proyección del gobierno de datos hacia una mayor cantidad de datos gobernados. En términos generales, el logro del objetivo principal tuvo un impacto positivo en la gestión y administración de la Universidad de Chile. En primer lugar, se logró inventariar y documentar los activos, procesos y flujos críticos de datos, lo que permitió obtener y democratizar información que nunca antes había estado disponible en la Universidad de Chile. Además, la implementación del plan de gobernanza mejoró la calidad de los datos en las tablas de nivel central, lo que a su vez permite proporcionar información validada a otras unidades que requieran información en el futuro.

A modo conclusivo, un punto importante a destacar dentro del desarrollo de la implementación del plan de gobernanza, fue la gran cantidad de problemas de calidad debido al ingreso de data en aplicaciones, donde si bien, tomando las acciones desarrolladas en la presente memoria se da un aseguramiento estratégico de entrada de datos hacia las tablas a gobernar, no se aborda la modificación de estas aplicaciones, situación que podría eventualmente perpetuar problemas de calidad hacia otras tablas a las que se genere conexión. De esta manera, la solución más factible para este problema es la modificación estratégica de estas aplicaciones, acción que incurriría en costos de priorización de trabajo y económicos, además de decisiones que deben ser tomadas por más áreas técnicas dentro de la universidad, escapando del accionar que tienen la Dirección de Datos.

En función de este problema, y como aprendizaje dentro del desarrollo de esta memoria, es claro ver como la proyección del plan de gobierno de datos debe ir evolucionando en función del tiempo, donde se debe poder converger el accionar de distintas áreas involucradas en los datos a gobernar, y así generar una transversalidad de decisión con base en el aseguramiento de la calidad de los datos.

Empíricamente el gobierno de datos ha constituido una práctica de gestión realmente trascendental para muchas organizaciones, y es así como en la Universidad de Chile encontró un espacio de acción importante para poder desplegar una base sólida que vaya proyectando en el tiempo la gobernabilidad de más atributos y tablas, convergiendo el accionar de áreas involucradas en los datos con la toma de decisiones a más alto nivel jerárquico dentro de la universidad. Finalmente, la naturaleza evolutiva del plan de gobernanza permitirá ir generando cada vez más una gobernabilidad robusta, entendiendo nuevas aristas de la gobernanza, y la gobernabilidad de más atributos indispensables para la gestión y administración de la institución.

## 10 TRABAJO FUTURO

En cuanto al trabajo realizado en la última etapa de la implementación para los atributos asociados a nombres y apellidos, y específicamente a la reparación de estos, los alcances se concentraron en poder limpiar los atributos que se diferenciaban en cuanto a tildes con lo encontrado en el registro civil, dejando los atributos por nombres y/o apellidos diferentes, y atributos que poseen *NULL* dentro de las bases transaccionales de la universidad fuera de la limpieza pactada para esta memoria. Por esta razón se recomienda poder abordar diferenciadamente los atributos que no cruzan por nombres y/o apellidos diferentes, y atributos que poseen *NULL* dentro de las bases transaccionales de la universidad, realizando reuniones con las áreas involucradas en este dato, y finalmente definir planes de acción en función de abordar la limpieza de estos registros.

El plan de gobierno de datos implementado en la Universidad de Chile posee características cíclicas, es decir, luego de poder gobernar los atributos pactados para esta memoria, se propone gobernar los demás atributos asociados a datos personales, en particular; fechas de nacimiento y fallecimiento, estado civil, sexo, país, ciudad, direcciones, entre otros, levantando documentación de flujos, identificación de roles y responsabilidades sobre esto, y en general toda la información que sea pertinente en función de los entregables para cada uno de los pasos, para así finalmente poder extenderse a más datos transaccionales de importancia para la institución. De esta manera, se recomienda que, dentro de estos atributos asociados a datos personales, se dé un especial énfasis en el atributo “SEXO”, en función de la emergente demanda que se ha generado para este atributo debido a la creación de la Dirección de Igualdad de Género, donde han solicitado de manera continua este atributo a la Dirección de Datos para desarrollar analíticas que aborden la igualdad de género dentro de la Universidad de Chile. La responsabilidad de asegurar el cumplimiento del plan y la gobernanza de los atributos correspondientes debe recaer en la Oficina de Gobernanza de Datos de la Universidad de Chile. Esta oficina tendrá un papel fundamental en el seguimiento del plan acordado y será clave en la implementación de los pasos necesarios. Sin embargo, es fundamental contar con una colaboración activa tanto de los comités estratégicos como de los técnicos, ya que el trabajo en equipo es esencial para avanzar en estos temas.

En cuanto al abordaje de la calidad de datos dentro de la gobernanza expresada por DAMA, la actual memoria presenta el modelado e implementación de un plan de gobernanza de datos enfocado en la gestión de calidad de datos, y desde este punto, se recomienda integrar nuevos indicadores de calidad definidos por el *framework* DAMA dentro del *dashboard* desplegado en PowerBI, en función de robustecer el análisis de la calidad sobre los atributos y tablas a gobernar. Se recomienda que la elección de estos indicadores de calidad sea en función de una priorización dada por las aristas más críticas que se posean, levantadas dada la realización de mesas de trabajo, para posteriormente ser aprobados por el comité estratégico de gobierno de datos.

Asimismo, se recomienda poder abarcar las demás aristas que contempla la gobernanza de datos dentro del *framework* DAMA, tales como gestión de integración de datos, seguridad de datos, modelado de datos, entre otras, y así robustecer la gobernanza dentro de todos los aspectos que componen la misma. Se recomienda que estas aristas sean debidamente integradas en los respectivos pasos ya definidos dentro de la metodología, o que se proyecten como la generación de nuevos pasos, mediante la creación de respectivas entradas y salidas que permitan materializar los respectivos entregables para cada uno de los pasos.

Por último, en cuanto a los alcances de fondo para la gobernanza de datos en la Universidad de Chile, se propone poder abordar un trabajo en conjunto con los distintos comités en cuanto a la definición de nuevas políticas y normativas asociadas a los datos a gobernar, las que permitan complementar nuevas acciones correctivas que favorezcan la gestión y administración de estos.

# 11 BIBLIOGRAFÍA

- (1) J. Effective Healthcare Data Governance: How One Hospital System is Managing its Data Assets to Improve Outcomes [En línea]. 2015. Disponible en: <[https://www.healthcatalyst.com/success\\_stories/data-governancein-healthcare](https://www.healthcatalyst.com/success_stories/data-governancein-healthcare)> [Consulta: 18 noviembre 2022]
- (2) PowerData. Consejos prácticos para el gobierno de datos: un estudio de caso [En línea]. 2015. Disponible en: <<https://blog.powerdata.es/el-valor-de-la-gestion-dedatos/bid/406200/consejos-prcticos-para-el-gobierno-de-datosun-estudio-de-caso>> [Consulta: 18 noviembre 2022]
- (3) Healthcatalyst. Improved Data Access Drives Effective Care Delivery. [En línea]. 2019. Disponible en: <[https://www.healthcatalyst.com/success\\_stories/analyticsadoption-texas-childrens-hospital](https://www.healthcatalyst.com/success_stories/analyticsadoption-texas-childrens-hospital)> [Consulta: 18 noviembre 2022]
- (4) Universidad de Chile: La Universidad [En línea]. 2022. Disponible en: <<https://www.uchile.cl/presentacion>> [Consulta: 15 Mayo 2022]
- (5) Eckerson, W, Data Quality and the Bottom line, TDWI, 2002
- (6) Friedman, T.; Smith, M, Measuring the Business Value of Data Quality, Gartner, 2011
- (7) VV. AA, Analytics, BI, and Information Management Survey, Information Week, 2011
- (8) DAMA Internacional. Guía de fundamentos para la gestión, chapter 1. Introducción, page 25. Technics Publications, LLC, 2009.
- (9) Power Data. Metadatos, definición y características. [En línea]. 2016. Disponible en: <https://www.powerdata.es/>. [Consulta: 20 Mayo 2022]
- (10) DAMA Internacional. Guía de fundamentos para la gestión, chapter 12. Gestión de calidad de datos, page 346. Technics Publications, LLC, 2009.
- (11) DAMA Internacional. Guía de fundamentos para la gestión, chapter 12. Gestión de calidad de datos, page 347. Technics Publications, LLC, 2009.
- (12) DAMA Internacional. Guía de fundamentos para la gestión, chapter 12. Gestión de calidad de datos, page 351. Technics Publications, LLC, 2009.
- (13) DAMA Internacional. Qué es DMBok [En línea]. 2022. Disponible en: <<https://www.damachile.org/dmbok>>. [Consulta: 16 Mayo 2022]
- (14) DAMA Internacional. Guía de fundamentos para la gestión, chapter 3. Gobierno de datos, page 63. Technics Publications, LLC, 2009.

- (15) The MDM Institute. What is Data Governance? [En línea]. 2016. Disponible en: <https://0046c64.netsolhost.com/whatIsDataGovernance.html>. [Consulta: 17 Mayo 2022]
- (16) Gwen Thomas. The DGI Data Governance DGI Data Governance Framework, chapter 1. What is Data Governance?, page 3. 2013.
- (17) IBM Corporation, Gobierno del ciclo de vida de la información [En línea]. 2016 Disponible en: <https://www.ibm.com/>. [Consulta: 17 Mayo 2022]
- (18) Sunil Soares. Data Governance with Oracle, chapter 1. Introduction, page 1. 2010
- (19) Oracle. What is database [En línea]. 2022. Disponible en: <https://www.oracle.com/cl/database/what-is-database/>. [Consulta: 18 Mayo 2022]
- (20) Power Data. Bases de datos transaccional, la mina de oro digital. [En línea]. 2016. Disponible en: <https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bases-de-datos-transaccionales-la-mina-de-oro-digital> [Consulta: 23 Mayo 2022]
- (21) Power Data. ¿Qué son los procesos ETL?. [En línea]. 2017. Disponible en: <https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/qu-son-los-procesos-etl> [Consulta: 23 Mayo 2022]
- (22) Google Cloud. Trabaja con procedimientos almacenados de SQL. [En línea]. 2022. Disponible en: <https://cloud.google.com/bigquery/docs/procedures?hl=es419#:~:text=Un%20procedimiento%20almacenado%20es%20un,conjunto%20de%20datos%20de%20BigQuery>. [Consulta: 23 Mayo 2022]
- (23) DAMA Internacional. Guía de fundamentos para la gestión, chapter 5. Desarrollo de datos page 142. Technics Publications, LLC, 2009.
- (24) RedHat. ¿ Qué es una API?. [En línea]. 2021. Disponible en: <https://www.redhat.com/es/>. [Consulta: 23 Julio 2022]
- (25) María Isabel Mejía Jaramillo. Guía Técnica - Gobierno del dato, chapter 1. Introducción, page 11. MINTIC, 2014.
- (26) Ministerio de Tecnologías de la Información y las Comunicaciones, Guía para la Gestión y Clasificación de Activos de Información. chapter 5. Definiciones, page 9-10, 2014.
- (27) María Isabel Mejía Jaramillo. Guía Técnica - Gobierno del dato, Glosario, page 45. MINTIC, 2014.
- (28) DAMA Internacional. Guía de fundamentos para la gestión, chapter 4. Gestión Arquitectura de Datos page 109. Technics Publications, LLC, 2009.

- (29) Microsoft. Access SQL: conceptos básicos, vocabulario y sintaxis. [En línea]. 2022. Disponible en: <https://support.microsoft.com/es-es/office/access-sql-conceptos-b%C3%A1sicos-vocabulario-y-sintaxis-444d0303-cde1-424e-9a74-e8dc3e460671>. [Consulta: 23 Julio 2022]
- (30) IBM. Lenguaje de consulta estructurada (SQL). [En línea]. 2022. Disponible en: <https://www.ibm.com/docs/es/db2woc?topic=reference-sql>. [Consulta: 24 Julio 2022]
- (31) OpenWebinars, ¿Qué es SQL server? [En línea]. 2022. Disponible en: <https://openwebinars.net/blog/que-es-sql-server/> >. [Consulta: 24 Julio 2022]
- (32) IONOS. Oracle Database: definición y funcionamiento. [En línea]. 2022. Disponible en: [ionos.es/digitalguide/hosting/cuestiones-tecnicas/oracle-database/#:~:text=Oracle%20Database%20es%20un%20sistema,estadounidense%20de%20software%20y%20hardware.>](https://ionos.es/digitalguide/hosting/cuestiones-tecnicas/oracle-database/#:~:text=Oracle%20Database%20es%20un%20sistema,estadounidense%20de%20software%20y%20hardware.>). [Consulta: 24 Julio 2022]
- (33) Analisisyprogramacionoop. Crear trabajos (Jobs) con SQL Server 2014. [En línea]. Disponible en: <https://analisisyprogramacionoop.blogspot.com/2016/01/trabajos-jobs-sql-server-2014.html>. [Consulta: 24 Julio 2022]
- (34) Microsoft. ¿Qué es Power BI?. [En línea]. 2022. Disponible en: <https://learn.microsoft.com/es-es/power-bi/fundamentals/power-bi-overview> . [Consulta: 28 Julio 2022]
- (35) Inmensia. Creación y Uso de DBLinks en Oracle. [En línea]. 2020. Disponible en: [https://inmensia.com/blog/20070329/database\\_link\\_oracle.html](https://inmensia.com/blog/20070329/database_link_oracle.html). [Consulta: 29 Julio 2022]
- (36) PowerData. Data Warehouse: todo lo que necesitas saber sobre almacenamiento de datos. [En línea]. 2022. Disponible en: <https://www.powerdata.es/data-warehouse/#:~:text=Un%20Data%20Warehouse%20es%20un,recuperar%20y%20f%C3%A1cil%20de%20administrar.>> . [Consulta: 24 Julio 2022]
- (37) DAMA Internacional. Guía de fundamentos para la gestión, chapter 9. Data Warehousing e inteligencias de negocios page 243. Technics Publications, LLC, 2009.
- (38) MSN. ¿Qué es y para qué sirve Visual Studio 2017?. [En línea]. 2022. Disponible en: <https://www.msn.com/es-cl/noticias/microsoftstore/%C2%BFqu%C3%A9-es-y-para-qu%C3%A9-sirve-visual-studio-2017/ar-AAAnLZL9#:~:text=Visual%20Studio%20es%20un%20conjunto,entorno%20que%20soporte%20la%20plataforma.>>. [Consulta: 12 Agosto 2022]
- (39) Danysoft. Visual Studio. [En línea]. 2022. Disponible en : <https://www.danysoft.com/visual-studio/>. [Consulta: 13 Agosto 2022]
- (40) DAMA Internacional. Guía de fundamentos para la gestión, chapter 7. Gestión de seguridad de datos page 202. Technics Publications, LLC, 2009.
- (41) IBM. Clarify roles and responsibilities by using a RACI matrix. [En línea]. 2004. Disponible en: <https://www.ibm.com/garage/method/practices/manage/raci-matrix/>. [Consulta: 20 Agosto 2022]

- (42) Microsoft. CREATE TRIGGER (Transact-SQL). [En línea]. 2022. Disponible en: <https://learn.microsoft.com/es-es/sql/t-sql/statements/create-trigger-transact-sql?view=sql-server-ver16> . [Consulta: 20 Agosto 2022]
- (43) Marco Villamar Coloma. Aplicación móvil para publicidad y ventas del almacén Babahoyo, capítulo 1. Marco Teórico page 18. 2017. <<https://dspace.uniandes.edu.ec/bitstream/123456789/8428/1/TUBMIE010-2017.pdf>>.
- (44) Rafael Menéndez-Barzanallana Asensio. JAVASCRIPT, parte 1. Conceptos básicos de Javascript page 2. UMU. <<https://dspace.uniandes.edu.ec/bitstream/123456789/8428/1/TUBMIE010-2017.pdf>>.
- (45) Gabriel Rodríguez Flores. Desarrollo de una aplicación web con Node.js para la monitorización en tiempo real de un electrocardiograma, Capítulo 4. Herramientas de desarrollo. page 26. Universidad de Sevilla. 2017 <[https://idus.us.es/bitstream/handle/11441/66716/TFG\\_Gabriel%20Rodr%C3%ADguez%20Flores.pdf?sequence=1&isAllowed=y](https://idus.us.es/bitstream/handle/11441/66716/TFG_Gabriel%20Rodr%C3%ADguez%20Flores.pdf?sequence=1&isAllowed=y)>.
- (46) Gabriel Rodríguez Flores. Desarrollo de una aplicación web con Node.js para la monitorización en tiempo real de un electrocardiograma, Capítulo 4. Herramientas de desarrollo. page 26. Universidad de Sevilla. 2017 <[https://idus.us.es/bitstream/handle/11441/66716/TFG\\_Gabriel%20Rodr%C3%ADguez%20Flores.pdf?sequence=1&isAllowed=y](https://idus.us.es/bitstream/handle/11441/66716/TFG_Gabriel%20Rodr%C3%ADguez%20Flores.pdf?sequence=1&isAllowed=y)>.
- (47) David Barrera Garay. Aplicaciones de software que permitan la integración con otras aplicaciones de otras plataformas en la empresa A, Capítulo 6. Referencia. page 95. Universidad tecnológica del Perú. 2020 <[https://repositorio.utp.edu.pe/bitstream/handle/20.500.12867/3543/David%20Barrera\\_Edgar%20Poma\\_Trabajo%20de%20Investigacion\\_Bachiller\\_2020.PDF?sequence=1&isAllowed=y](https://repositorio.utp.edu.pe/bitstream/handle/20.500.12867/3543/David%20Barrera_Edgar%20Poma_Trabajo%20de%20Investigacion_Bachiller_2020.PDF?sequence=1&isAllowed=y)>.
- (48) DAMA Internacional. Guía de fundamentos para la gestión, chapter 1. Introducción, page 30. Technics Publications, LLC, 2009.
- (49) DAMA Internacional. Guía de fundamentos para la gestión, chapter 1. Introducción, page 65. Technics Publications, LLC, 2009.
- (50) MinTic. Acerca del MinTIC. [En línea]. 2017. Disponible en: <https://www.mintic.gov.co/portal/inicio/Ministerio/Acerca-del-MinTIC/>. [Consulta: 16 septiembre 2022]
- (51) María Isabel Mejía Jaramillo. Guía Técnica - Gobierno del dato, chapter 1. Introducción, page 11. MINTIC, 2014.

# ANEXOS



Maria Magdalena Gandolfo  
Firmado digitalmente por Maria Magdalena Gandolfo Gandolfo  
Fecha: 2022.06.29 15:41:52 -04'00'

## CREA COMITÉ ESTRATÉGICO DE GOBIERNO DE DATOS EN LA UNIVERSIDAD DE CHILE Y FIJA SUS NORMAS DE ORGANIZACIÓN Y FUNCIONAMIENTO.

DECRETO EXENTO N°0018675

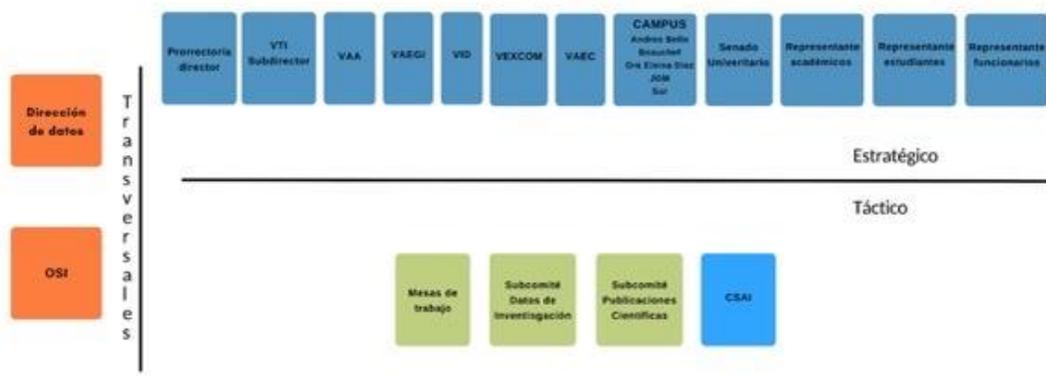
SANTIAGO, 16 de junio de 2022.

### VISTOS:

Lo dispuesto en el D.F.L. N°3, de 2006, que fija el texto refundido, coordinado y sistematizado del D.F.L. N°153, de 1981, que establecen los Estatutos de la Universidad de Chile, ambos del Ministerio de Educación; el Decreto Supremo N°199, de 2018, del referido Ministerio; la Ley 21.094, sobre universidades estatales; la Ley N°19.880, que establece Bases de los Procedimientos Administrativos que rigen los actos de los Órganos de la Administración del Estado; el D.U. N°0022604, de 2020, que crea la Vicerrectoría de Tecnologías de la Información; el Decreto TRA N°309/56/2021; el D.U. N°0044208, de 2017; la Resolución N°7, de 2019, de la Contraloría General de la República

### Anexo A Decreto creación comité estratégico

#### Conformación comité estratégico



### Anexo B Estructura de los comités

Estructura mesa de datos



Anexo C Estructura mesa de datos

Mesas de trabajo Gobierno de datos

	Nombres-Apellidos		RUT		Direcciones	
	Áreas involucradas	Representante en la mesa	Áreas involucradas	Representante en la mesa	Áreas involucradas	Representante en la mesa
	Nombres áreas y representantes	Pregrado	Carlos Rilling	Pregrado	Carlos Rilling	Pregrado
Postgrado		Veronica Zuniga	Postgrado	Veronica Zuniga	Postgrado	Veronica Zuniga
Personal		Pilar Labarca	Personal	Pilar Labarca	Personal	Pilar Labarca
VTI-Mantenición		Jorge Ibarra/Julio Huidobro	VTI-Mantenición	Jorge Ibarra/Julio Huidobro	VTI-Mantenición	Jorge Ibarra/Julio Huidobro
VTI-Desarrollo		Manuel Alba	VTI-Desarrollo	Manuel Alba	VTI-Desarrollo	Manuel Alba
Ucampus		Victor Hernandez	Ucampus	Victor Hernandez	Ucampus	Victor Hernandez
TyG		Maria Teresa Gomez	TyG	Maria Teresa Gomez	TyG	Maria Teresa Gomez
Aranceles		Andrea Santoro	Aranceles	Andrea Santoro	Aranceles	Andrea Santoro
AUGE		Ernestina Guzmán	AUGE	Ernestina Guzmán	AUGE	Ernestina Guzmán
DGP		Fabiola Divin	DGP	Fabiola Divin	DGP	Fabiola Divin
DBE		Ann Fleet /Rodrigo Carmona	DBE	Ann Fleet /Rodrigo Carmona	DBE	Ann Fleet /Rodrigo Carmona
DEMRE		Pamela Becerra	DEMRE	Pamela Becerra	DEMRE	Pamela Becerra
Proyectos TI (VTI)		José Miguel Munoz	Proyectos TI (VTI)	José Miguel Munoz	Proyectos TI (VTI)	José Miguel Munoz
ADMDB (VTI)		Rodrigo Diaz	ADMDB (VTI)	Rodrigo Diaz	ADMDB (VTI)	Rodrigo Diaz
Sistemas (VTI)	Sandra Jaque/Nelson Lopez	Sistemas (VTI)	Sandra Jaque/Nelson Lopez	Sistemas (VTI)	Sandra Jaque/Nelson Lopez	

Anexo D Estructura mesa de datos



UNIVERSIDAD DE CHILE

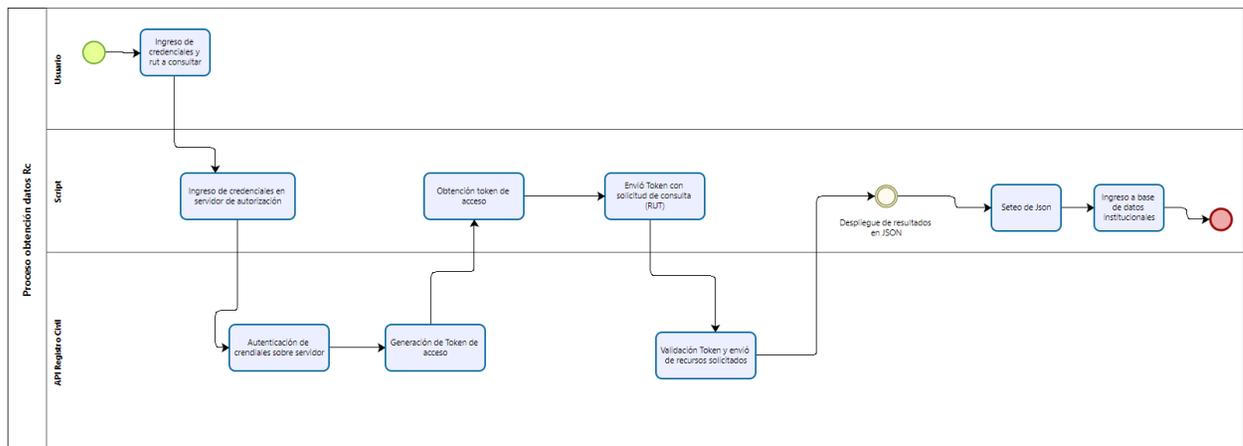
UNIVERSIDAD DE CHILE

## Calidad de datos

*Autor:*  
Dirección de Datos

25 de abril de 2022

### *Anexo E Documento estandarización datos*



### *Anexo F Proceso carga API Registro Civil*

```

var sec = require("../core/sec");
var axios = require('axios');
var querystring = require('querystring');
var db = require('oracledb');
//var log = require("../core/log");

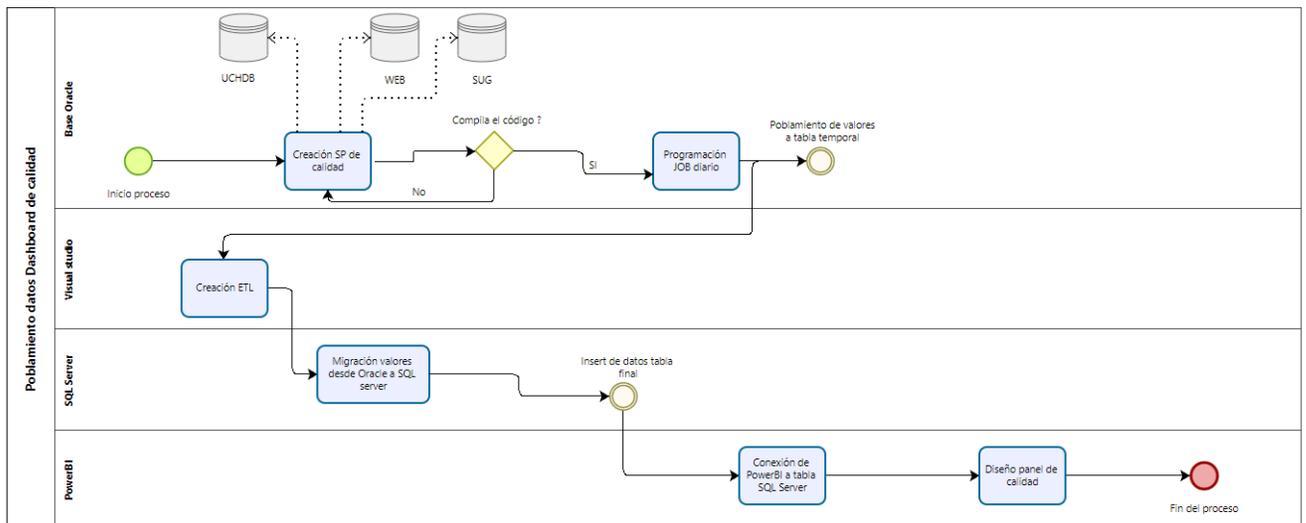
let ruta_api = "/rut/:rut";

let dbconfig = {
  user: "APP_SYNC_SRCEI",
  password: "y53W13c0ShcIYCn",
  connectString: "(DESCRIPTION=(ADDRESS=(PROTOCOL=TCP)(HOST=orauchp-cluster-scan.uchile.cl)(PORT=1525))(CONNECTOR=(PROTOCOL=OCI)"));
};

function setPersona(rut, nombres, paterno, materno, f_nacimiento, f_defuncion, estado_civil, genero, nacionalidad) {
  return {
    rut: rut,
    nombres: nombres,
    paterno: paterno,
    materno: materno,
    primario: paterno,
    secundario: materno,
    fechaNacimiento: f_nacimiento,
    fechaDefuncion: f_defuncion?f_defuncion:"",
    estadoCivil: estado_civil,
    genero: genero,
    nacionalidad: nacionalidad,
    fallecido: fallecido
  };
};

```

*Anexo G Script de carga API Registro Civil*



*Anexo H Proceso de creación Dashboard de Calidad*

```

Código Detalles Referencias Dependencias Errores Perfiles Permisos
SUGINTEGRACION

create or replace PROCEDURE PRC_DASHBOARD_CALIDAD
/*AUTOR: Marco Olguin Decencier*/
IS
BEGIN

EXECUTE IMMEDIATE 'TRUNCATE TABLE INTEGRACION_UCHDB.TMP_INDICADOR';

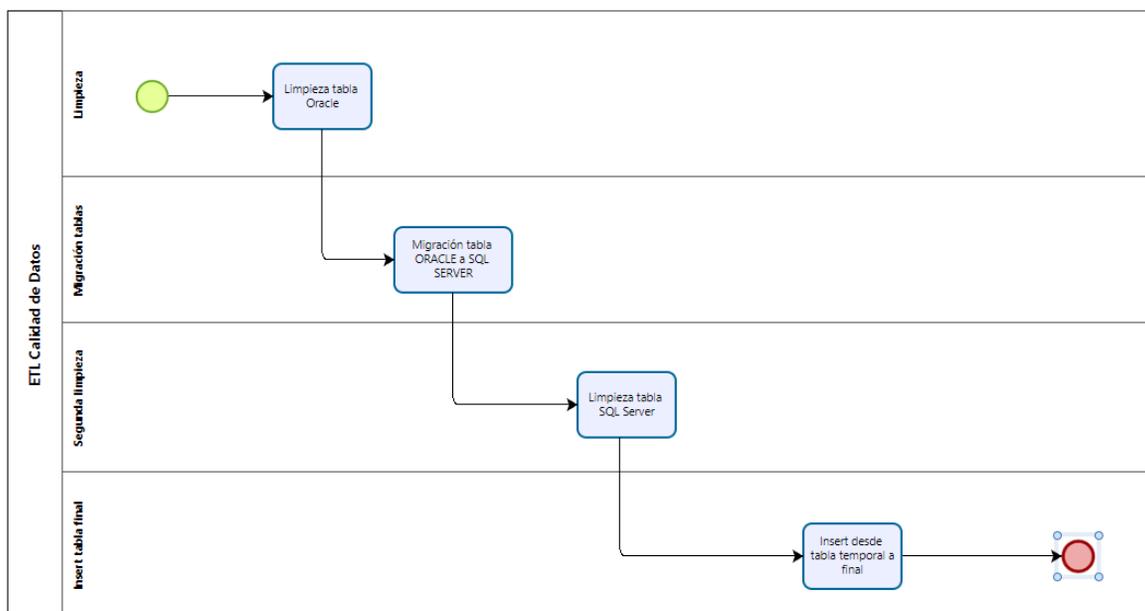
COMMIT;

--- >>>> CICLO EXACTITUD INDIVIDUO >>>>>>

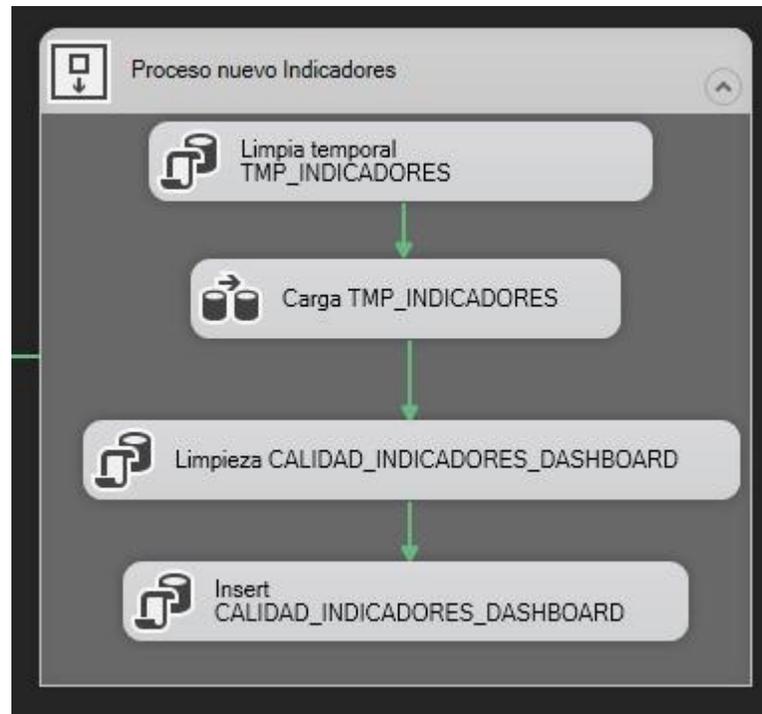
INSERT INTO INTEGRACION_UCHDB.TMP_INDICADOR SELECT NOCUMPLE, CUMPLE, TOTAL, 'NOMBRES' AS ATRIBUTO, 'INDIVIDUO' AS FUENTE, 'EXACTITUD' AS INDICADOR,
FROM (
    select count(1) as cumple from
    (
        select rut as Rutc, NOMBRES, NOMBRE_RC FROM INTEGRACION_UCHDB.CRUCES_OPERACIONAL_REGISTRO WHERE FUENTE = 'Individuo (alumnos)'
    )
    where NOMBRES = NOMBRE_RC
) t1
cross join
(
    select count(1) as nocumple from
    (
        select rut rutnocum, NOMBRES, NOMBRE_RC FROM INTEGRACION_UCHDB.CRUCES_OPERACIONAL_REGISTRO WHERE FUENTE = 'Individuo (alumnos)'
    )
    where NOMBRES <> NOMBRE_RC
) t2
cross join
(
    select count(1) as total FROM INTEGRACION_UCHDB.CRUCES_OPERACIONAL_REGISTRO WHERE FUENTE = 'Individuo (alumnos)'
)

```

Anexo I Procedimiento almacenado



Anexo J BPMN ETL



### Anexo K ETL en Visual Studio

```

create or replace FUNCTION    fx_reemplaza_may_min (ls_frase IN VARCHAR2)
RETURN VARCHAR2
IS
  ls_frase_dos  VARCHAR2 (500) := NULL;
  ln_pos        NUMBER;
  ln_paso       number(1);
  --x varchar2(50);
BEGIN
  IF ls_frase IS NULL OR ls_frase = ' '
  THEN
    ls_frase_dos := ls_frase;
  ELSE
    ls_frase_dos := INITCAP (ls_frase);

    IF ls_frase in ('De Garrido', 'De Gregorio', 'Del Barrio', 'De Latorre', 'Samuel En Lai', 'Del Mar', 'La Vega', 'DEL DESPÓSITO', 'LA TORRE')
    THEN
      ln_paso:=0;
    ELSE
      ls_frase_dos := REPLACE (ls_frase_dos, 'De ', 'de ');
      ls_frase_dos := REPLACE (ls_frase_dos, 'de Luca', 'De Luca');
      ls_frase_dos := REPLACE (ls_frase_dos, 'Del ', 'del ');
      ls_frase_dos := REPLACE (ls_frase_dos, 'La ', 'la ');
      ls_frase_dos := REPLACE (ls_frase_dos, 'Y ', 'y ');
      ls_frase_dos := REPLACE (ls_frase_dos, 'En ', 'en ');
      ls_frase_dos := REPLACE (ls_frase_dos, 'Con ', 'con ');
      ls_frase_dos := REPLACE (ls_frase_dos, 'Y ', 'y ');
      ls_frase_dos := REPLACE (ls_frase_dos, 'Los ', 'los ');
      ls_frase_dos := REPLACE (ls_frase_dos, 'Uno ', 'uno ');
      ln_pos := INSTR (ls_frase_dos, 'Dos');
    END IF;

    IF ln_pos = 1
    THEN
      IF ls_frase_dos = 'Dos Santos' OR ls_frase_dos = 'Dos Reis'
      THEN
        ls_frase_dos := ls_frase_dos;
      ELSE
        ls_frase_dos := REPLACE (ls_frase_dos, 'Dos ', 'dos ');
      END IF;
    END IF;
  END IF;

```

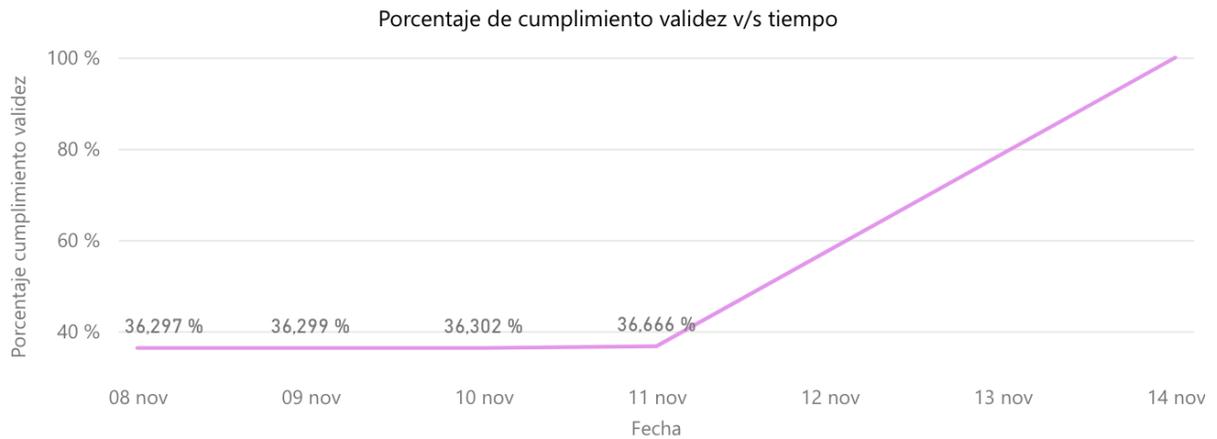
```

ls_frase_dos := REPLACE (ls_frase_dos, 'Tres ', 'tres ');
ls_frase_dos := REPLACE (ls_frase_dos, 'Cuatro ', 'cuatro ');
ls_frase_dos := REPLACE (ls_frase_dos, 'Cinco ', 'cinco ');
ls_frase_dos := REPLACE (ls_frase_dos, 'Seis ', 'seis ');
ls_frase_dos := REPLACE (ls_frase_dos, 'Siete ', 'siete ');
ls_frase_dos := REPLACE (ls_frase_dos, 'Ocho ', 'ocho ');
ls_frase_dos := REPLACE (ls_frase_dos, 'Nueve ', 'nueve ');
ls_frase_dos := REPLACE (ls_frase_dos, 'Puntos ', 'puntos ');
ls_frase_dos := REPLACE (ls_frase_dos, 'Min:', 'min:');
ls_frase_dos := REPLACE (ls_frase_dos, 'Max:', 'max:');
ls_frase_dos := REPLACE (ls_frase_dos, 'Mención ', 'mención ');
ls_frase_dos := REPLACE (ls_frase_dos, 'Mención ', 'mención ');
ls_frase_dos := REPLACE (ls_frase_dos, '"', '');
ls_frase_dos := REPLACE (ls_frase_dos, 'E ', 'e ');
ls_frase_dos := REPLACE (ls_frase_dos, 'Ha ', 'ha ');
ls_frase_dos := REPLACE (ls_frase_dos, 'El ', 'el ');
ls_frase_dos := REPLACE (ls_frase_dos, 'Grado ', 'grado ');
ls_frase_dos := REPLACE (ls_frase_dos, 'Por ', 'por ');
ls_frase_dos := REPLACE (ls_frase_dos, 'Para ', 'para ');
ls_frase_dos := REPLACE (ls_frase_dos, 'Otorgado ', 'otorgado ');
ls_frase_dos := REPLACE (ls_frase_dos, 'Of ', 'of ');
ls_frase_dos := REPLACE (ls_frase_dos, 'Revalidado ', 'revalidado ');
ls_frase_dos := REPLACE (ls_frase_dos, 'Reconocido ', 'reconocido ');
ls_frase_dos := REPLACE (ls_frase_dos, 'El ', 'el ');
ls_frase_dos := REPLACE (ls_frase_dos, 'Carácter ', 'carácter ');
ls_frase_dos := REPLACE (ls_frase_dos, 'Terminal', 'terminal');
ls_frase_dos := REPLACE (ls_frase_dos, 'la Habana', 'La Habana');
ls_frase_dos := REPLACE (ls_frase_dos, 'Como', 'como');
ls_frase_dos := REPLACE (ls_frase_dos, 'Asimilable', 'asimilable');
ls_frase_dos := REPLACE (ls_frase_dos, 'Al ', 'al ');
ls_frase_dos := REPLACE (ls_frase_dos, 'Título ', 'título ');
ls_frase_dos := REPLACE (ls_frase_dos, 'Diploma ', 'diploma ');
ls_frase_dos := REPLACE (ls_frase_dos, 'Las ', 'las ');
ls_frase_dos := REPLACE (ls_frase_dos, '(Mba)', '(MBA)');
ls_frase_dos := REPLACE (ls_frase_dos, 'las Heras', 'Las Heras');
END IF;

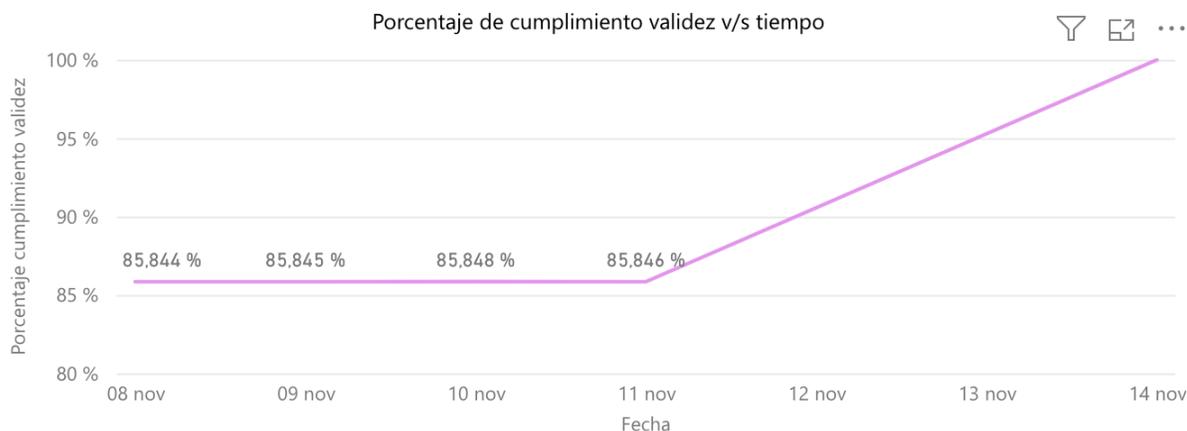
RETURN TRIM (ls_frase_dos);
END;

```

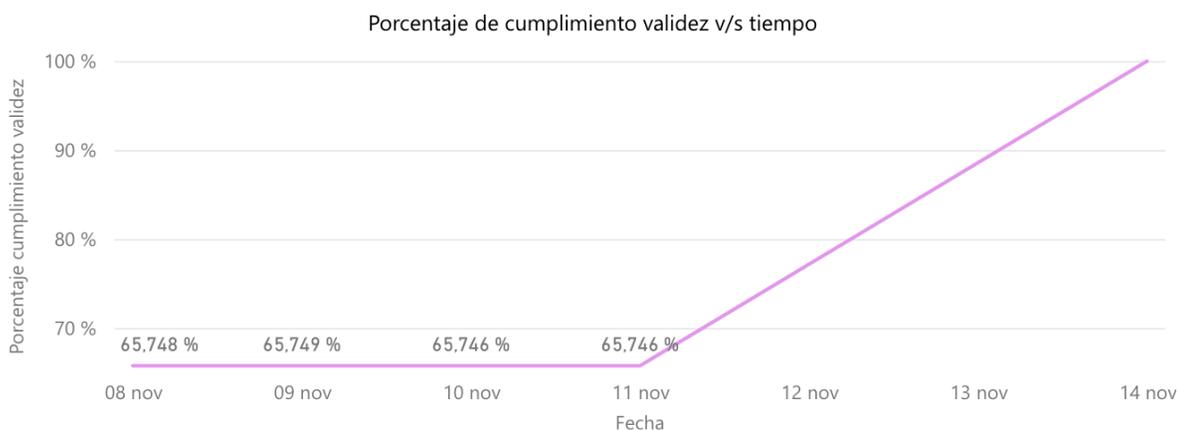
### Anexo L Función de formateo



### Anexo M Resultados indicador validez (INDIVIDUO)



*Anexo N Resultados indicador validez (DOTACION)*



*Anexo O Resultados indicador validez (DATOS PERSONALES)*

```

create or replace PROCEDURE          IND_NOMBRES_FIX_TILDE
IS
BEGIN
  DECLARE V_CNT INT;
  BEGIN
    Loop
      EXECUTE IMMEDIATE ('TRUNCATE TABLE DATAGOB.RUT_TEXT');

      INSERT INTO DATAGOB.RUT_TEXT(RUT, TEXTO)
      SELECT RUT, IND_NOMBRES_FIX
      FROM DATAGOB.LISTADO_CAMBIO_NOMBRE_TILDES L
      WHERE IND_NOMBRES_FIX IS NOT NULL
      AND ACTUALIZADO_IND_NOMBRE = 0
      AND ROWNUM<10000;
      DBMS_OUTPUT.PUT_LINE('INSERT EN BUFFER: ' || SQL%ROWCOUNT || ' REGISTROS CARGADOS.');
```

```

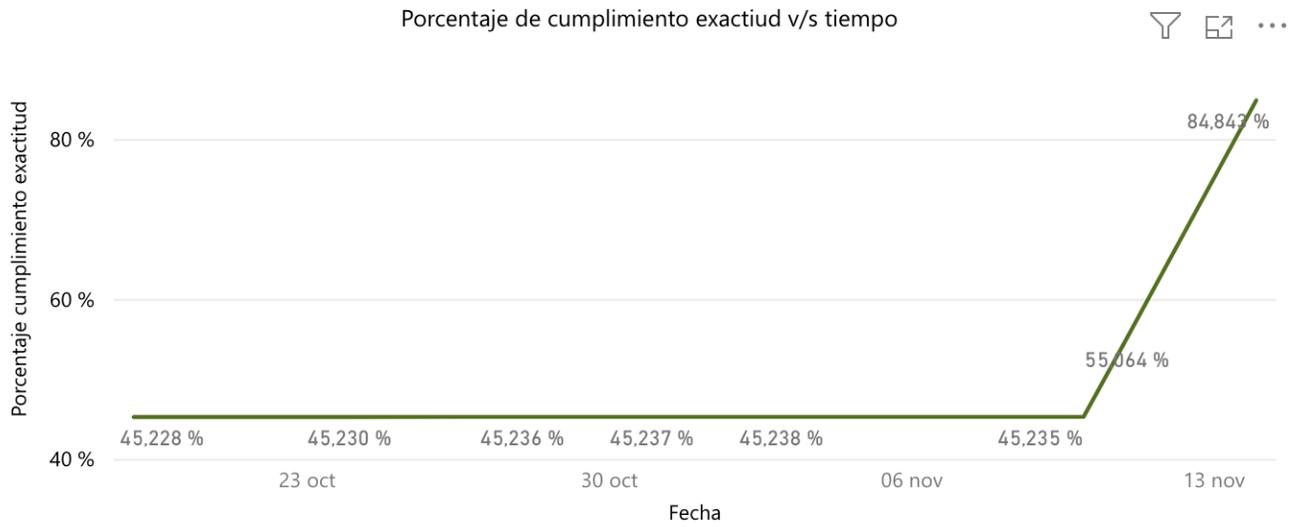
      MERGE INTO SUG.INDIVIDUO I
      USING (SELECT RUT, TEXTO FROM DATAGOB.RUT_TEXT) L
      ON (I.ID = L.RUT)
      WHEN MATCHED THEN
      UPDATE
      SET NOMBRES = L.TEXTO;
      V_CNT := SQL%ROWCOUNT;
      DBMS_OUTPUT.PUT_LINE(V_CNT || ' REGISTROS ACTUALIZADOS.');
```

```

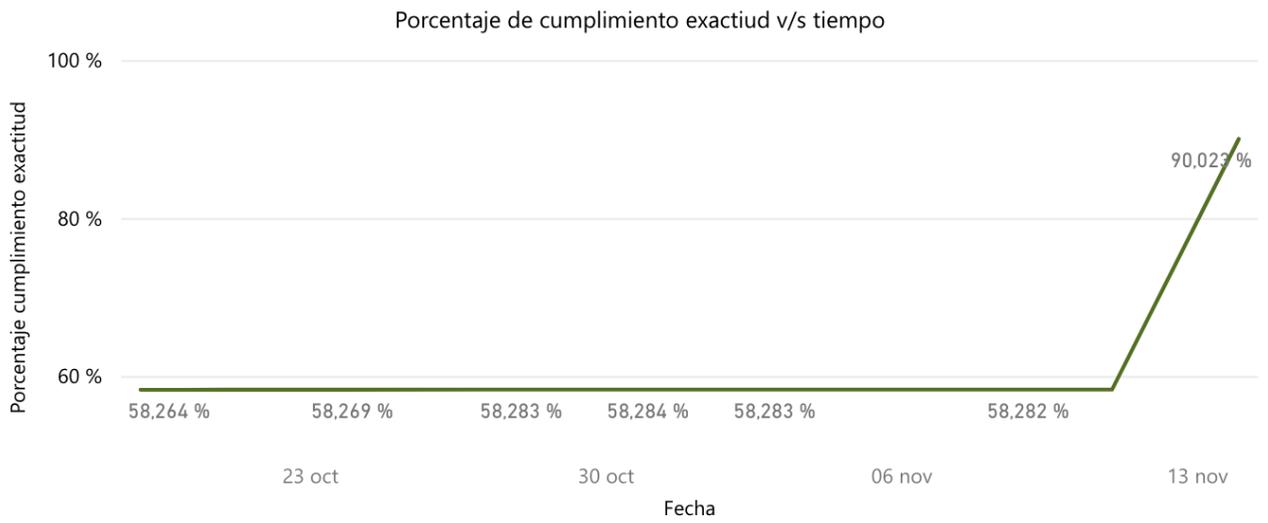
  END LOOP;

```

*Anexo P Procedimiento almacenado de migración*

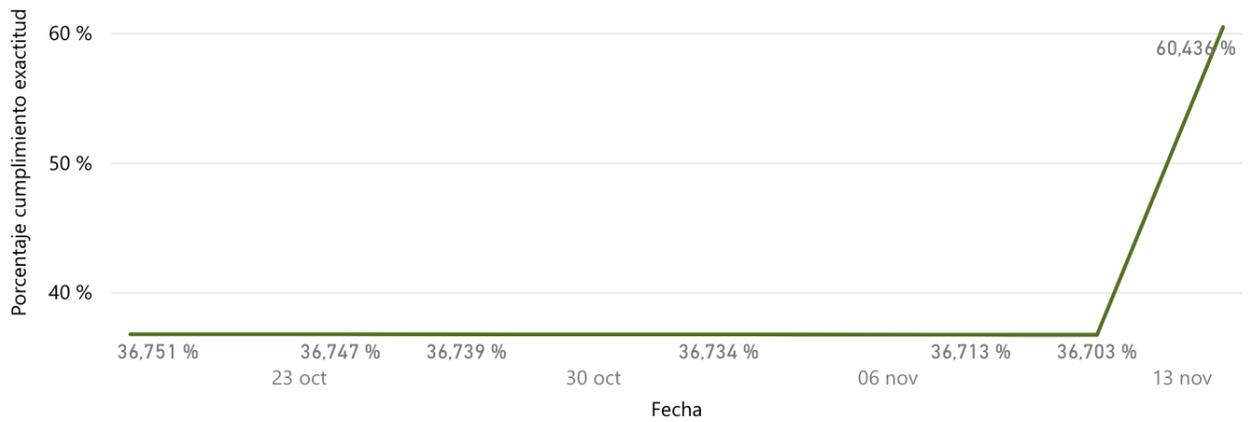


*Anexo Q Porcentaje de cumplimiento exactitud v/s tiempo (INDIVIDUO)*



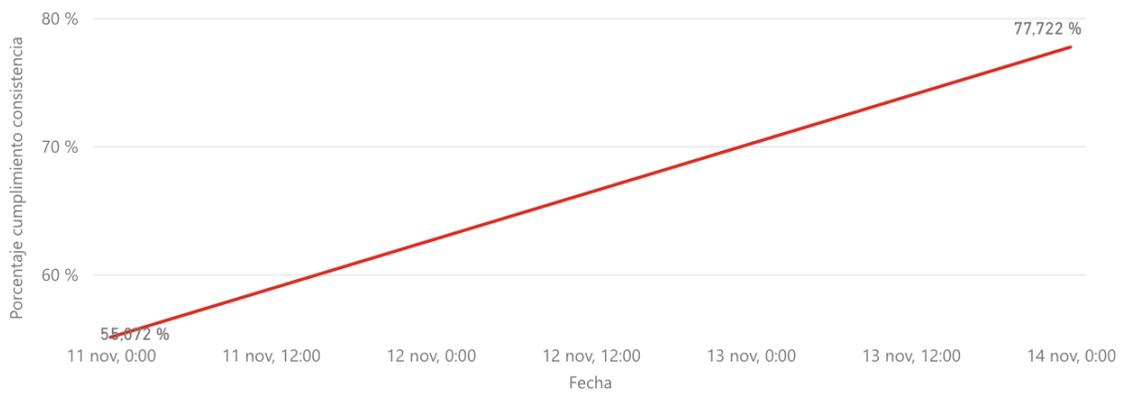
*Anexo R Porcentaje de cumplimiento exactitud v/s tiempo (DOTACION)*

Porcentaje de cumplimiento exactitud v/s tiempo



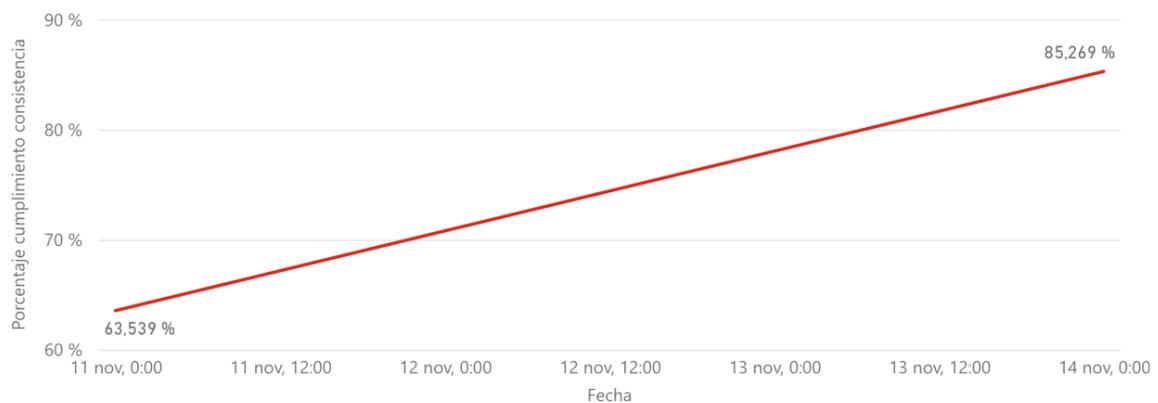
**Anexo S Porcentaje de cumplimiento exactitud v/s tiempo (DATOS PERSONALES)**

Porcentaje de cumplimiento consistencia v/s tiempo

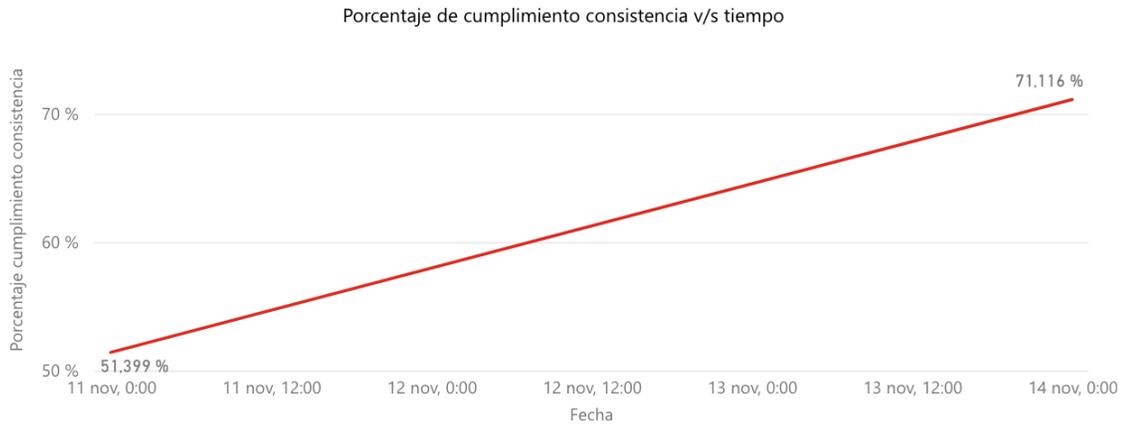


**Anexo T Porcentaje de cumplimiento consistencia v/s tiempo (INDIVIDUO)**

Porcentaje de cumplimiento consistencia v/s tiempo



**Anexo U Porcentaje de cumplimiento consistencia v/s tiempo (DOTACION)**



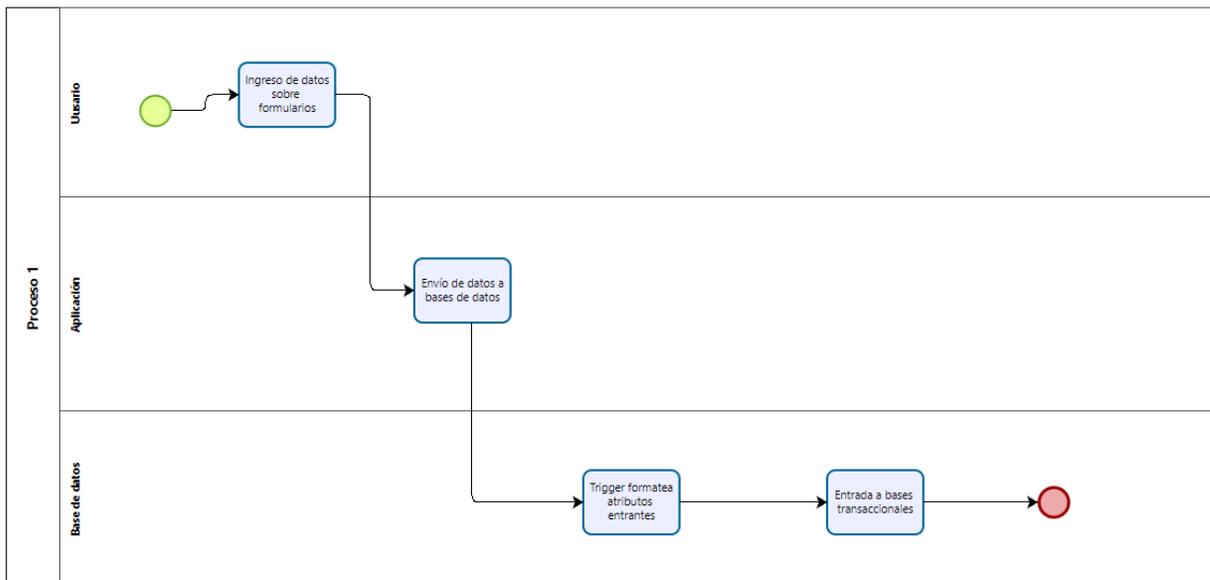
*Anexo V Porcentaje de cumplimiento consistencia v/s tiempo (DATOS PERSONALES)*

```

CREATE OR REPLACE TRIGGER SUG.PERSONA_NOMBRE
BEFORE
  INSERT OR UPDATE
ON SUG.INDIVIDUO
REFERENCING NEW AS NEW OLD AS OLD
FOR EACH ROW
BEGIN
  SELECT
    SUG.fx_reemplaza_may_min(:NEW.NOMBRES),
    SUG.fx_reemplaza_may_min(:NEW.PATERNO),
    SUG.fx_reemplaza_may_min(:NEW.MATERNO)
  INTO :NEW.NOMBRES, :NEW.PATERNO, :NEW.MATERNO
  FROM DUAL;
END;

```

*Anexo W Trigger validez*



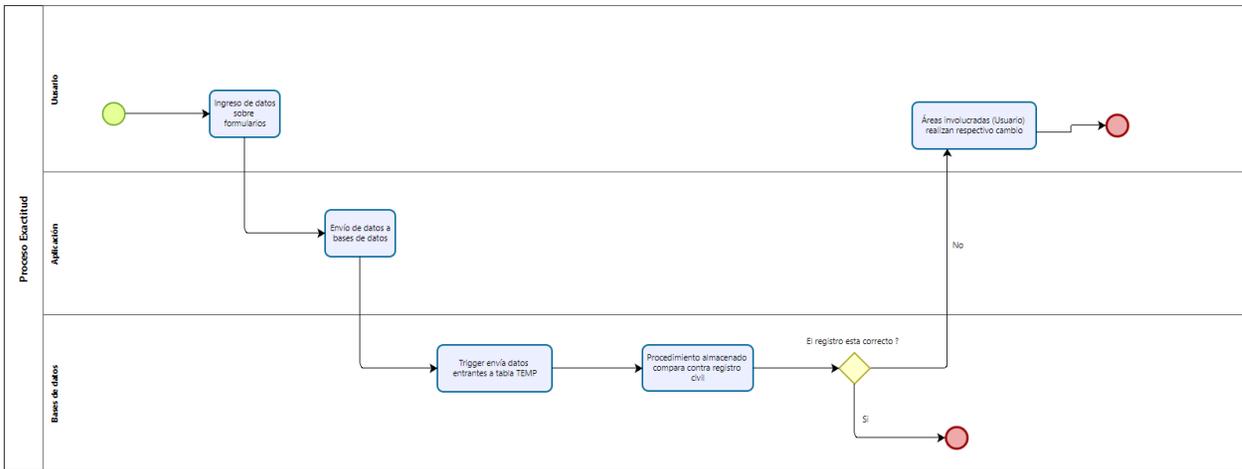
*Anexo X Proceso de limpieza data entrante (INDIVIDUO)*

```

CREATE OR REPLACE TRIGGER SUG.TRG_GOB_EXACTITUD_PERSONAS
AFTER INSERT OR UPDATE
OF ID,
PATERNO
,MATERNO
,NOMBRES
,SEXO
,FECHA_NACIM
,ESCI_ID
ON SUG.INDIVIDUO
REFERENCING NEW AS New OLD AS Old
FOR EACH ROW
DISABLE
WHEN (
NVL(UPPER(new.paterno), '') <> NVL(UPPER(old.paterno), '')
OR NVL(UPPER(new.materno), '') <> NVL(UPPER(old.materno), '')
OR NVL(UPPER(new.nombres), '') <> NVL(UPPER(old.nombres), '')
OR NVL(UPPER(new.sexo), '') <> NVL(UPPER(old.sexo), '')
OR NVL(UPPER(new.esci_id), '') <> NVL(UPPER(old.esci_id), '')
OR (TRUNC(new.fecha_nacim) <> TRUNC(old.fecha_nacim) OR old.fecha_nacim IS NULL OR new.fecha_nacim IS NULL)
)
)
DECLARE
/*****

```

*Anexo Y Trigger exactitud*



*Anexo Z Proceso de validación data entrante*