

PROGRAMA DE MODERNIZACIÓN DEL SECTOR PÚBLICO

**ESTUDIO DE USO INTENSIVO DE DATOS EN
POLÍTICAS PÚBLICAS: ANÁLISIS Y
RECOMENDACIONES ESTRATÉGICAS PARA LA
IMPLEMENTACIÓN DE UNA POLÍTICA EN BASE A LA
EVIDENCIA INTERNACIONAL**

DICIEMBRE DE 2017

PATRICIO RODRÍGUEZ

CONSULTORÍA REALIZADA ENTRE
21 NOVIEMBRE Y 31 DICIEMBRE 2017

CONTRAPARTE TÉCNICA

FELIPE GONZÁLEZ

JOSÉ INOSTROZA

JUAN ANDRÉS ROESCHMANN

RANDALL LEDERMANN

SERVICIOS ESPECIALIZADOS N°

851556-1-IN1



CONTENIDO

1	INTRODUCCIÓN	5
1.1	PROBLEMA	5
1.2	LA EVIDENCIA COMO UN ACTIVO ESTRATÉGICO PARA EL SECTOR PÚBLICO.....	6
1.3	EL ANÁLISIS DE DATOS PARA LA GENERACIÓN DE EVIDENCIA.....	8
1.4	EJEMPLOS DE USO DE DATOS EN EL SECTOR PÚBLICO EN CHILE Y EL MUNDO	10
2	MARCO CONCEPTUAL	15
2.1	LA CADENA DE GENERACIÓN DE VALOR A PARTIR DE LOS DATOS	15
2.2	ECOSISTEMA DE GENERACIÓN DE VALOR A PARTIR DE LOS DATOS.....	16
2.3	MODELO DE MADUREZ DE GOBIERNO BASADO EN DATOS.....	19
2.4	MODELO DE TRANSFORMACIÓN DIGITAL EN EL SECTOR PÚBLICO	20
3	DIAGNÓSTICO.....	21
3.1	PUBLICAR: DISPONIBILIZAR DATOS PARA LA CIUDADANÍA Y LAS INSTITUCIONES PÚBLICAS.	21
3.2	REFINAR: ASEGURAR LA CALIDAD DE LOS DATOS PÚBLICOS	24
3.3	ANALIZAR: EXPLOTAR EL CONOCIMIENTO CONTENIDO EN LOS DATOS PÚBLICOS	26
3.4	OPTIMIZAR: ANÁLISIS AVANZADO DE DATOS COMO POLÍTICA DE ESTADO.....	27
3.5	DIAGNÓSTICO A TRAVÉS DEL MODELO DE TRANSFORMACIÓN DIGITAL.....	28
4	ESTRATEGIA PARA LA IMPLEMENTACIÓN EFECTIVA DE UN GOBIERNO BASADO EN DATOS..	29
4.1	DEFINICIONES CLAVE DE LA ESTRATEGIA	30
4.2	COMPONENTES DEL SISTEMA DE GOBIERNO BASADO EN DATOS.....	31
4.3	ESPECIFICACIONES DE LA ESTRATEGIA	34
4.4	INICIATIVAS PRIORITARIAS EN EL CORTO Y MEDIANO PLAZO.....	39
4.5	ESTRATEGIA A TRAVÉS DEL MODELO DE TRANSFORMACIÓN DIGITAL.....	40
	REFERENCIAS	42
	ANEXO: EJEMPLOS DE METODOLOGÍAS ASOCIADAS A LA GESTIÓN DE DATOS Y SERVICIOS TECNOLÓGICOS AFINES	48

Uso intensivo de datos en políticas públicas: análisis y recomendaciones estratégicas para la implementación de una política en base a la evidencia internacional: análisis y recomendaciones estratégicas

Patricio Rodríguez ^a, Felipe González-Zapata ^b, José Inostroza ^b,
Juan Andrés Roeschmann ^c, Randall Ledermann ^b

^a Centro de Investigación Avanzada en Educación, Universidad de Chile, Periodista José Carrasco
Tapia 75, Santiago, Chile.

^b Programa de Modernización del Sector Público, Ministerio de Hacienda, Teatinos 120, Piso 11,
Santiago, Chile.

^c Dirección de Presupuestos, Ministerio de Hacienda, Teatinos 120, Piso 12, Santiago, Chile.

Resumen Ejecutivo — El análisis de datos ha mostrado ser una herramienta efectiva tanto para mejorar la gestión del Estado como para detectar desafíos y oportunidades a partir de la evidencia para invertir en una serie de programas que pueden enfrentar los desafíos y problemas futuros. Sin embargo, actualmente nuestro país no está aprovechando esta oportunidad, principalmente por la falta de implementación de una gobernanza que permita desarrollar una cadena de generación de valor a partir de los datos en el sector público, esto es cómo a partir del dato generado al interior del Estado, este se recolecta, se procesa, almacena, se comparte con otras reparticiones y finalmente se transforme en evidencia que alimente la toma de decisiones y el diseño, evaluación, implementación y seguimiento de políticas públicas.

El presente documento analiza la situación actual de Chile respecto a las políticas de datos abiertos y uso intensivo de datos en las instituciones públicas, y propone una **estrategia para desarrollar un *data-driven government* al año 2022**, basado en dos componentes principales. La primera componente es un *sistema de producción, recolección y disponibilización de datos*, de responsabilidad de la División de Gobierno Digital (DGD). Esta entidad bajo el liderazgo de un *Chief Data Officer*, se encargará de implementar la gobernanza de la cadena de generación de valor a partir de datos, coordinando la intencionalidad y priorización de publicación de datos abiertos que alimenten la participación de la ciudadanía y el sector privado promoviendo un ecosistema de innovación abierta.

La segunda componente, es el desarrollo de analítica en entidades claves que usen significativamente los datos para la optimización de la toma de decisiones, y el diseño, evaluación e implementación de políticas pública, para producir casos de éxito temprano. Una de dichas entidades es que requiere instalar una *unidad de analítica es la Dirección de Presupuesto* del Ministerio de Hacienda que debe tomar decisiones anuales presupuestarias del orden de 65.000 millones de dólares y diseña sistemas de largo plazo.

Se propone que este proceso sea acompañado por organismos internacionales de reconocido prestigio como la OCDE y BID, desarrollando también relaciones de cooperación con otros países que hayan recorrido un camino importante en este ámbito como son Nueva Zelanda, Francia y el Reino Unido, y universidades y centros de investigación tanto nacionales como internacionales.

1 INTRODUCCIÓN

1.1 PROBLEMA

En los últimos 30 años, Chile ha experimentado un notable crecimiento económico. En 1990, el producto interno bruto per cápita de nuestro país era de USD **4.589**, quintuplicándose 26 años después alcanzando los USD **23.960** (World Bank Group, 2017). A pesar de ese notable crecimiento económico, nuestro país es una de las naciones con más alta desigualdad de los países pertenecientes a la Organización para la Cooperación y el Desarrollo Económico (OCDE), con un índice GINI de 0,454 (2015) siendo solo superada por México (OECD, 2017d).

Por lo tanto, en un mundo globalizado y competitivo en el que nuestro país se ha integrado a través de numerosos acuerdos de libre comercio, nuestra nación enfrenta una serie de desafíos de diversificación de su economía, matriz energética, desarrollo de capital humano y migratorios, entre otros, que requieren de un Estado moderno, ágil en la toma de decisiones y asignación de recursos, y rápido y eficiente en la implementación de políticas públicas proactivas. Sin embargo, existe el diagnóstico generalizado que el Estado presenta una brecha importante para alcanzar estos objetivos debido a, entre otras razones, la falta de análisis avanzado de datos en su políticas públicas (Centro de Estudios Públicos, 2017).

La falta de sistemas de análisis avanzado de datos hace que el Estado esté perdiendo tanto oportunidades como recursos por falta de un adecuado diseño y ejecución de políticas públicas basadas en evidencia, y en muchos casos donde lamentablemente el Estado ha sido ineficaz para encontrar soluciones. El Servicio Nacional de Menores¹ (SENAME), es un ejemplo elocuente. Entre enero de 2005 y el 30 de junio de 2016, **1.313 personas** fallecieron estando bajo la protección del Estado ya sea en centros de su dependencia directa o en instituciones colaboradoras (Radio Villa Francia, 2016). En particular para el caso de esta institución, hay una falta significativa de datos para generar el conocimiento adecuado que permita a las autoridades tomar mejores decisiones y anticiparse a la ocurrencia de este tipo de problemas.

Adicionalmente existen diversas oportunidades de mejora que el Estado pierde porque no es capaz de identificarlas y accionarlas. Un ejemplo en este ámbito es el del gasto en educación. A modo de ejemplo: si todos los años existen estudiantes que repiten y otro que abandonan el sistema escolar, entonces ¿cuál es la magnitud del costo financiero de este problema, y cómo podría reducirse?

Por ejemplo, usando como fuente los datos de matrícula publicados en formato abierto por el Ministerio de Educación (Ministerio de Educación, 2017), se estimó cuál es el gasto adicional para el Estado de pagar las subvenciones escolares nuevamente para aquellos estudiantes que repiten un curso, especialmente más de una vez y para los alumnos que abandonan para las cohortes de estudiantes que ingresaron a primer básico entre los años 2004 y 2015 inclusive (Rodríguez & Pinto, 2017). A modo de síntesis, el costo directo de la exclusión educativa (deserción más repetencia), tuvo **al menos** un costo actualizado para el país de **631.313 millones**² de pesos. Si se considera que el

¹ El SENAME es un organismo gubernamental centralizado, colaborador del sistema judicial y dependiente del Ministerio de Justicia, que encarga de la protección de derechos de niños, niñas y adolescentes, y de jóvenes entre 14 y 18 años ante el sistema judicial. Además, se ocupa de regular y controlar la adopción en Chile.

² Tomando el valor de la UF del 24 de octubre de 2017, que es de CLP 26.646,94.

fenómeno se mantendrá, en base al comportamiento de los últimos tres años, el **valor anual** promedio de recursos perdidos es **79.085** millones² de pesos.

De no tomarse medidas para reducir las tasas actuales de repitencia y deserción, el valor actual de la pérdida a perpetuidad, considerando una tasa social de descuento³ del 6%, ascenderá a **1.318.085 millones²** de pesos. Este valor corresponde aproximadamente al **0,9%** del PIB (2016) de Chile⁴. Esta evidencia no considera el valor presente de la pérdida de ingresos debido a la reducción de la vida laboral de cada persona producto de la repitencia, la diferencia de ingresos producto de un menor nivel de escolaridad, y otras pérdidas para el Estado como mayor inversión en subsidios en Salud y pensiones. Un estudio reciente estima el valor presente de dicho costo en un 2,6% del PIB de 2011 (Kokkidou, Rodríguez, & Mondaca, 2017).

Los casos presentados ilustran, aunque sea anecdóticamente, algunas de las cuantiosas pérdidas que tiene el Estado por no hacer una gestión basada en datos. Una gestión que permita acercarnos al Estado moderno, ágil y eficiente que el país requiere. El Estado chileno es un país de ingresos medios que, por una parte, requiere de más recursos para invertir en buenos programas de inversión social y que, por otra parte, pierde recursos que podrían ser justamente la fuente de financiamiento de dichos programas. Este desafío de diseñar y monitorear las políticas públicas en base a evidencia ha sido objeto de múltiples estudios, en donde por ejemplo se estimó que una mayor eficiencia fiscal en el gobierno de EE.UU. podría generar un ahorro potencial de 17 billones de dólares (Harvard Kennedy School, 2017; Wiseman, 2017b). Como consecuencia, la gestión pública moderna debe basarse en evidencia; ***“Gobernar sin datos es como manejar un auto con los ojos vendados”*** (Curtatone, 2014; Wiseman, 2016).

El Presidente electo Sebastián Piñera afirmó durante su campaña su intención en ahorrar alrededor de USD 7.000 millones para financiar su programa de gobierno. Este desafío, además de detectar oportunidades de inversión social que permitan reducir gastos en el futuro, requiere de información para determinar en qué, cuándo y en cuánto se puede aumentar la eficiencia del gasto fiscal, por ejemplo en ahorros operacionales, mejor toma de decisiones y reasignación de recursos: **¿Con qué datos y capacidades analíticas cuenta el Estado para guiar sus procesos de toma de decisiones, y el diseño, desarrollo, implementación y evaluación de actuales y nuevas políticas públicas? ¿Cómo podemos hacer más eficiente la gestión del estado usando analítica de datos?**

1.2 LA EVIDENCIA COMO UN ACTIVO ESTRATÉGICO PARA EL SECTOR PÚBLICO

Gracias al desarrollo de las tecnologías de información, los datos han pasado de ser escasos a ser superabundantes (Manyika et al., 2011; Staff Science, 2011; The Economist, 2010). Como otros factores esenciales para la producción y la prestación de servicios (por ejemplo activos fijos y capital humano), gran parte de la gestión pública moderna necesita de datos como insumo para su desarrollo, lo que representa al mismo tiempo desafíos y oportunidades para las instituciones públicas respecto al proceso de gestión y uso de estos datos (Manyika et al., 2011; Staff Science, 2011). Dado que actualmente el volumen de datos que se genera sobrepasa la capacidad de almacenarlos físicamente (Hilbert & López, 2011), el reto para el sector público es gestionarlos de

³ Según Ministerio de Desarrollo Social.

⁴ Fuente: Banco Central de Chile, tomando un tipo de cambio al dólar Observado al 30 de diciembre de 2016.

forma adecuada para generar información y, finalmente, conocimiento de valor público, incluso más allá de uso primario. Es por ello que los datos representan un **valioso activo** para los actores públicos y la comunidad cívica, dado que pueden ser re-usados para apoyar la toma de decisiones basadas en evidencia, la transparencia y rendición de cuentas, y la mejora en la entrega de servicios entre distintos sectores públicos (Ubaldi, 2013; Gonzalez-Zapata & Heeks, 2015).

El uso de datos para tomar decisiones tiene una larga trayectoria en el sector privado, donde las empresas usan cálculos complejos sobre los datos de sus clientes y sus procesos operacionales usando técnicas denominadas de “Inteligencia de negocios” (del inglés *“Business intelligence”*) para descubrir patrones y tendencias que predigan el comportamiento futuro de los consumidores y operaciones (Bienkowski, Feng, & Means, 2012), mejorando la efectividad de sus estrategias. Existe evidencia significativa que el uso de datos para la toma de decisiones tiene un impacto importante en el desarrollo de las economías de cada país y sus ciudadanos, creando valor para la economía mundial, mejorando la productividad y competitividad privada y pública, y generando excedentes económicos para los consumidores (Manyika et al., 2011).

En particular, el objetivo de la analítica es generar evidencia que sea pertinente, de calidad y oportuna, para así fundamentar y orientar decisiones estratégicas en base a la evidencia, lo que no necesariamente ocurre de forma frecuente en las instituciones públicas. Esto significa diagnosticar problemas que pasan inadvertidos o desapercibidos y que son, por lo tanto, imposibles de accionar (Bienkowski et al., 2012). Este proceso se denomina “toma de decisiones basada en evidencia” (del inglés *‘Data-driven decision making’*) (Provost & Fawcett, 2013). El análisis de datos también puede ser utilizado para la mejora de la gestión pública, a través de la generación de más y soluciones más precisas que satisfagan necesidades de salud, educación, transporte, vivienda, atención e inclusión de grupos desaventajados, entre otras, a partir de contextos sociales, demográficos y territoriales particulares (Rodríguez, Palomino, & Mondaca, 2017).

Parte importante del rol que la analítica ha cobrado en el sector público se debe al énfasis que los gobiernos, las organizaciones de la sociedad civil y diversas instituciones internacionales han puesto en el acceso abierto a los datos que generan y administran las instituciones públicas. Las iniciativas de **datos abiertos** (del inglés *‘open data’*) consisten en la apertura de los datos públicos en formatos legibles por máquinas y bajo licencias abiertas que permitan su reutilización para cualquier propósito (Open Knowledge International, 2017). Específicamente, los datos abiertos facilitan el análisis de datos porque reducen los costos de transacción de datos entre las instituciones públicas, al mismo tiempo que permiten que agentes fuera del sector público (tales como organizaciones de la sociedad civil, universidades, centros de investigación, o *think tanks*, entre otros) realicen análisis avanzado de datos utilizando estas fuentes, lo que a su vez puede generar beneficios burocráticos, políticos, tecnológicos y económicos (Gonzalez-Zapata & Heeks, 2015). Actualmente, la publicación de datos abiertos es una estrategia efectiva para la creación de valor público a través del análisis de estos datos tanto dentro como fuera de las instituciones públicas.

De esta manera, con el acceso a datos masivos y el uso de técnicas analíticas adecuadas, se pueden identificar y medir problemáticas que habían permanecido invisibles, lo que, obviamente, las hacía imposibles de gestionar. De esta forma, es posible desarrollar una “inteligencia de valor público” (un equivalente social de la “inteligencia de negocios”), que tiene la potencialidad de ser un componente estratégico para la toma de decisiones y el diseño, implementación y evaluación de políticas públicas (Rodríguez et al., 2017).

1.3 EL ANÁLISIS DE DATOS PARA LA GENERACIÓN DE EVIDENCIA⁵

Los datos tienen poco valor por sí mismos, ya que éstos deben ser analizados para la generación de conocimiento y valor público (Gandomi & Haider, 2015). Para el procesamiento y análisis de datos ha surgido una disciplina denominada **Ciencia de Datos** (Provost & Fawcett, 2013; Song & Zhu, 2016) (del inglés '*Data Science*'), que combina un conjunto amplio de técnicas provenientes de las Ciencias de la Computación, Matemáticas, Estadística, Econometría e Investigación Operativa (Marsh, Pane, & Hamilton, 2006; UNESCO, 2012). El ciclo de vida del análisis de datos considera a lo menos seis pasos, que se detallan en la Figura 1.

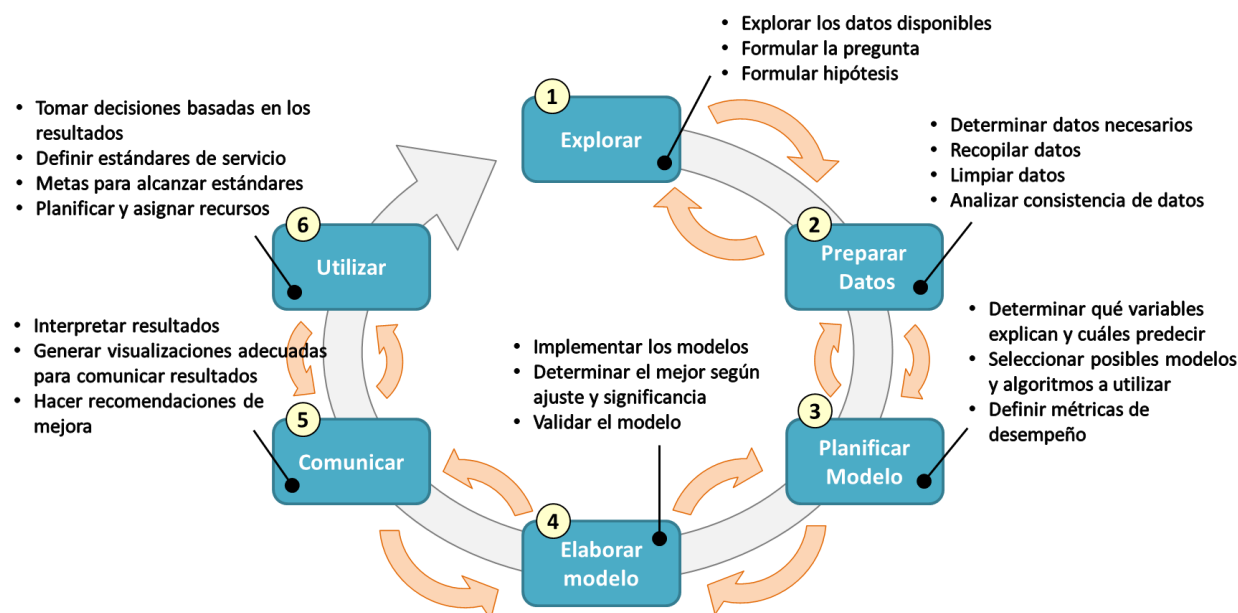


Figura 1: Ciclo de vida del análisis de datos (Schmarzo, 2013).

Como muestra la Figura 1, el ciclo de vida del análisis de datos no es lineal; muchas veces se requiere definir y reformular preguntas en función de la disponibilidad de datos, o reinterpretar los resultados a la luz de nueva evidencia. Por ende, es un proceso iterativo en que se puede retroceder a etapas previas. De todas maneras, se puede describir el procesamiento de los datos masivos a partir de dos etapas principales: la gestión de los datos y la analítica de datos (Gandomi & Haider, 2015; Labrinidis & Jagadish, 2012).

La **gestión de los datos** se compone de tres tareas: (1) adquisición y almacenamiento de los datos, (2) limpieza y depuración de los datos y, (3) la preparación para su análisis. Para dicha gestión, existe una amplia gama de metodologías que permiten realizar las tareas descritas previamente cuando se trata de fuentes de información no tradicionales y masivas (ver Tabla A1 en Anexo).

La **analítica de datos**, por otra parte, se refiere a la respuesta de preguntas y/o hipótesis formuladas a partir de técnicas de modelamiento y análisis. Como se puede constatar, este proceso no es particularmente distinto del proceso de indagación científica presente en cualquier disciplina; la

⁵ Adaptado de Rodríguez et al. (2017)

principal diferencia subyace en las características generales de los datos que se utilizan cuando son masivos (del inglés *'Big Data'*) con sus respectivos desafíos de acceso y manipulación de los datos. Para ello, existen distintas metodologías, esto es, los métodos científicos que se utilizan para el análisis (Tabla 1) a través de diferentes implementaciones tecnológicas en forma de productos y servicios.

Tabla 1: Ejemplos de metodologías para modelar y analizar datos (Adaptado de Rodríguez *et al.* (Rodríguez *et al.*, 2017)).

Metodología	Descripción	Aplicaciones / Ejemplos
Análisis espacial	Conjunto de técnicas que analizan las propiedades geométricas, topológicas y geográficas de un conjunto de datos (Joseph & Bantock, 1982; Luo & Qi, 2009; Luo & Wang, 2003; Radke & Mu, 2000; Wei, 2013).	<ul style="list-style-type: none"> • Regresiones espaciales (consumo vs. distancia a centros comerciales), simulaciones (desempeño de una cadena de distribución con bodegas en distintos lugares).
Análisis de redes	Conjunto de técnicas que caracterizan relaciones entre nodos discretos en un grafo o una red (Cormen, Leiserson, Rivest, & Stein, 2009).	<ul style="list-style-type: none"> • Identificación de líderes de opinión para focalizar campañas de marketing. • Identificar cuellos de botella en flujos de información de una empresa. • Modelamiento de redes de transporte y predicción del tiempo de desplazamiento de un punto a otro.
Aprendizaje automático (Machine Learning, ML)	<p>Subespecialidad de la Ciencia de la Computación (denominada históricamente "Inteligencia Artificial") que se ocupa del diseño y desarrollo de algoritmos que permiten inferir comportamientos basados en datos empíricos. El aprendizaje automático puede ser de dos tipos: supervisado y sin supervisión.</p> <p>En el aprendizaje supervisado se debe inferir una función a partir de un conjunto de ejemplos de entrenamiento. Estos consisten en un conjunto de entradas (en forma de vector) y un conjunto de salidas que son casos exitosos (satisfacen la función). Los casos exitosos permiten generar una medida de error respecto a las predicciones que se quieren hacer.</p> <p>En el caso del aprendizaje no supervisado estos "casos exitosos" no se conocen (o no se necesitan) y, por lo tanto, no existe retroalimentación para ajustar una función. El objetivo del algoritmo es organizar los datos o describir su estructura.</p>	<ul style="list-style-type: none"> • Predicción de fenómenos como crimen, deserción escolar y universitaria, esperanza de vida post-operatoria, ventas. • Sugerencias y recomendaciones de productos en función de historial pasado. • Procesamiento de lenguaje natural: reconocimiento de voz y lenguaje para interacción humano computador (ej: Siri, Cortana, Alexa), y análisis de sentimientos en textos y redes sociales. • Reconocimiento de patrones: texto manuscrito, procesamiento de imágenes y reconocimiento de caras para buscar sospechosos de crímenes. • Detección de anomalías: detección de fraudes bancarios en base a actividad inusual en compras usando tarjeta de crédito.
Inteligencia Territorial	Metodologías de análisis espacial que, a través de tecnologías de información, combinan enfoques cualitativos, cuantitativos y espaciales, respetando además los enfoques de participación y aproximación global multidisciplinaria y multisectorial. Por ejemplo, análisis de agrupamiento y valores atípicos espaciales (Anselin, 1995; Miller, 2004) en conjunto con análisis multicriterio que combinan diferentes características espaciales (Barredo, Kasanko, McCormick, & Lavalle, 2003).	<ul style="list-style-type: none"> • Indicadores espaciales de nivel de servicios públicos y privados. • Análisis de brechas y crecimiento espacial de la oferta y la demanda de servicios. • Accesibilidad urbana y rural, caracterización de los territorios en diferentes dimensiones geográficas, sociodemográficas, económicas y desarrollo humano.
Optimización	Técnicas numéricas de modelamiento para rediseñar y mejorar tanto procesos como sistemas complejos en múltiples dimensiones.	<ul style="list-style-type: none"> • Localización óptima de recursos en hospitales, escuelas, centros productivos, bodegas. • Producción: programación de maquinarias para la fabricación, manejo de inventarios.

Metodología	Descripción	Aplicaciones / Ejemplos
Pruebas A/B	Una técnica en la que un grupo de control se compara con una variedad de grupos de prueba con el fin de determinar qué cambios mejorarán una variable objetivo dada. Por esto, esta técnica también se conoce como <i>Split testing</i> o <i>Buckettesting</i> . El gran volumen de datos permite ejecutar y analizar un gran número de pruebas, asegurando que los grupos son de tamaño suficiente para detectar diferencias estadísticamente significativas entre los grupos de control y tratamiento).	<ul style="list-style-type: none"> • Probar la efectividad de distintas campañas de marketing. • Probar la efectividad de un tratamiento médico, o un tipo de educación a través de experimentos naturales donde algunos sujetos participan de una intervención y otros no, por diferentes circunstancias. Se busca además que estos sujetos sean lo más parecido en sus características (pareo) para controlar la mayor cantidad de variables posible.
Simulación	Modelación del comportamiento de sistemas complejo para pronóstico, predicción y planificación de escenarios.	<ul style="list-style-type: none"> • Pronóstico de los resultados financieros de una empresa dado circunstancias de incertidumbre. • Pronóstico del clima.
Visualización analítica de datos	Forma de descubrir y entender patrones en grandes conjuntos de datos vía interpretación visual, para que así los usuarios pueden navegar y explorar los datos.	<ul style="list-style-type: none"> • Análisis visual interactivo de componentes principales (Bienkowski et al., 2012; Jolliffe, 2002).
Visualización de datos	Comunicación de información en forma clara y efectiva a través de distintas formas de representación gráfica interactiva (Friedman, 2008; Lima, 2011; Steele & Iliinsky, 2010; Yau, 2011).	<ul style="list-style-type: none"> • Infografías. • Tableros de mando (<i>Dashboards</i>), para seguimiento y síntesis de ciertos fenómenos.

Finalmente hay diversos productos y servicios tecnológicos (software tanto bajo licencia abierta como propietaria) que permiten gestionar y analizar los datos, algunos de los cuales se ejemplifican en la Tabla A2 del Anexo A.

Este ciclo de análisis que supone la Ciencia de Datos requiere de profesionales especialistas con una formación sólida en alguna de las ciencias de la computación, uso y desarrollo de aplicaciones, modelamiento, estadística, analítica y matemáticas. A estos profesionales se les denomina **Científicos de Datos** (del inglés *Data Scientists*), quienes exploran, generan preguntas, realizan análisis de escenarios (“¿qué pasa si?”) y cuestionan los supuestos y procesos existentes utilizando múltiples fuentes de datos de diferentes orígenes (IBM, 2015; Song & Zhu, 2016).

1.4 EJEMPLOS DE USO DE DATOS EN EL SECTOR PÚBLICO EN CHILE Y EL MUNDO

Para visualizar más concretamente las oportunidades que el país está perdiendo por no incorporar el uso de analítica de datos como parte íntegra de su operación, mostramos una serie de casos que se han implementado tanto en Chile como en otros países que ejemplifican cómo puede mejorarse la gestión del Estado.

AGRICULTURA

En Colombia, el sistema AEPS (Agricultura Específica por Sitio) permitió determinar las condiciones favorables y desfavorables para las plantaciones dependiendo de su etapa de crecimiento (las que están variando debido al cambio climático) a través de técnicas de análisis de *Machine Learning* (ML) y utilizando datos de cosechas comerciales y clima (World Bank Group, 2016). Con esta información, se aconsejó a los agricultores las fechas óptimas de siembra y cosecha dependiendo del tipo de plantación, mejorando sus ingresos y compensando el conocimiento ya obsoleto debido a la variación del clima.

DESARROLLO SOCIAL

- **Detección de fraudes en ayuda y servicios:** los programas de ayuda y servicios sociales sufren fraudes y abusos que ponen en peligro el apoyo a las personas más necesitadas. (Goldsmith, Crawford, & Weinryb Grohsgal, 2016). Debido al enorme costo de implementar auditorías externas y detección manual del fraude, el gobierno local de Los Ángeles en EE.UU implementó un sistema de detección automática de fraude que identificó 200 casos en el año 2012, con un ahorro de 6,8 millones de dólares (Goldsmith et al., 2016).
- **Identificación de la pobreza:** En Sri Lanka, a través del uso de imágenes satelitales para identificar bienes e infraestructura de los hogares (por ejemplo, áreas construidas, sombras, tipos de techo y calles, entre otros) en combinación con información censal, fue posible dar cuenta de diferencias regionales en los niveles de pobreza en forma bastante precisa (World Bank Group, 2016). Esto permitió generar indicadores de pobreza, tanto en Sri Lanka como Pakistán, de forma más económica y exacta (World Bank Group, 2016). Esfuerzos similares se han realizado en Malawi, Nigeria, Tanzania y Uganda (Jean et al., 2016). Estas técnicas, en conjunto con el análisis de datos de llamadas de teléfonos móviles y de información administrativa han permitido identificar grupos objetivos de la política pública a una fracción del costo de las encuestas o censos, tal como ocurrió en Guatemala (Hernández, Hong, Frias-Martinez, Whitby, & Frias-Martinez, 2017).
- **Servicios sociales:** La Fundación para la Juventud de Medway (EE.UU.) usa análisis predictivo para identificar adultos jóvenes en situación de riesgo de puedan presentar problemas en el futuro. Con esta información, se les entrega a estos jóvenes los servicios y oportunidades que necesitan para tener éxito y se les orienta para que no tomen decisiones que resulten tanto peores para ellos y más costosas para el Estado en el largo plazo (Goldsmith et al., 2016). Por otro lado, el Departamento de servicios para personas sin hogar de la ciudad de Nueva York, junto con la fundación *SumAll*, han estudiado los datos de registro de entrada a los refugios, con el propósito de determinar las causas de la necesidad de abrigo y centrar los recursos en prevenir la falta de vivienda, más que reaccionar a sus consecuencias (Goldsmith et al., 2016).

EDUCACIÓN

- **Deserción escolar:** Una importante aplicación del análisis predictivo en la educación es detectar a estudiantes en riesgo de abandonar la escuela (Goldsmith et al., 2016). En los países de la OECD se calcula que uno de cada tres adultos tiene educación secundaria incompleta (OECD, 2010). De esta proporción, solo el 50% tiene empleo regular versus el 74% de los que si completaron sus estudios (OECD, 2010). Un aspecto relevante a estimar en base a esta información es el costo económico asociado a la deserción escolar secundaria; estimaciones propias sitúan esta cifra para Chile en el 2,6% del PIB (Kokkidou et al., 2017), mientras que en EE.UU alcanza el 1,6% del PIB (Rouse, 2005). Para predecir éste u otros indicadores deserción escolar se han utilizado distintas técnicas estadísticas y de ML (Cortez & Silva, 2008; Microsoft, 2016; Ministerio de Educación, 2015). En EE.UU, en el Distrito de Tacoma se implementó un modelo predictivo para determinar el nivel de deserción escolar en la educación pública de esta ciudad (Cortana Intelligence and ML Blog Team, 2015). En 2007, de los 30.000 estudiantes en educación secundaria, solamente el 55% de ellos lograba graduarse a tiempo, comparado con el 81% del promedio nacional. Para el 2014, focalizando las intervenciones escolares usando los resultados del modelo predictivo, se logró subir la tasa de graduación a tiempo de los estudiantes a un 78% (Cortana Intelligence and ML

Blog Team, 2015). En Chile, se han desarrollado distintos algoritmos para el mismo propósito (Cortana Intelligence and ML Blog Team, 2015; Cortez & Silva, 2008; Escobar & Lolas, 2015; Microsoft, 2016; Ministerio de Educación, 2015), pero que todavía no son parte de alguna política pública en la materia.

- **Deserción universitaria:** En Colombia, El Ministerio de Educación Nacional (MEN), en conjunto con las Instituciones de Educación Superior (IES), implementaron un proceso nacional para encontrar causas de deserción estudiantil a nivel de Programa, Institución Educativa o a nivel agregado: Sistema de Prevención y Análisis de la Deserción en las Instituciones de Educación Superior (SPADIES) (Centro de Estudios sobre Desarrollo Económico, 2014). Usando como base la estimación del riesgo de deserción estudiantil usando modelos estadísticos, las IES implementan metodologías y/o estrategias para disminuir el porcentaje de personas que abandonan sus estudios (Centro de Estudios sobre Desarrollo Económico, 2014). Técnicas predictivas similares también se aplican en otros países (Dekker, Pechenizkiy, & Vleeshouwers, 2009; Kalles, Pierrakeas, & Xenos, 2008; Kovacic, 2012; Nghe, Janecek, & Haddawy, 2007; Pal, 2012; Thai-Nghe, Horváth, & Schmidt-Thieme, 2011). Por ejemplo, en India se desarrolló un modelo para predecir la deserción de estudiantes de primer año, donde se obtuvo una exactitud del 91,7% en la predicción de los casos en donde los estudiantes no desertan y un 69,7% para los casos que si desertaron. En los Países Bajos, se creó un modelo para predecir la deserción de estudiantes de primer año de la carrera Ingeniería Eléctrica de la *Eindhoven University of Technology*, logrando un 80% de precisión en la identificación de los alumnos afectados por exclusión educativa (Dekker et al., 2009).
- **Gestión de la educación pública:** en Chile, para apoyar la puesta en marcha de los nuevos servicios locales de educación pública (SLEP) que se harán cargo progresivamente de la gestión de los establecimientos públicos en lugar de los municipios (Ministerio de Educación, 2017), el Laboratorio de Analítica Pública⁶ desarrolló una plataforma territorial que visualiza (por medio de mapas) indicadores de inequidad geográfica sociodemográfica y educativa que constituyen un diagnóstico de brechas y sirven de línea base para diseñar, implementar y monitorear los planes de acción de los SLEP y coordinar acciones con otras instituciones del Estado para los próximos años, según la nueva ley (Rodríguez, 2017).

FISCALIZACIÓN

En Chile, la Dirección del Trabajo está desarrollando modelos predictivos para mejorar la eficacia de las inspecciones de dicha dirección con el apoyo del Programa de Modernización del Sector Público. Resultados preliminares muestran una mejora de la eficacia de la fiscalización desde un 26,7% a un 42,3%. Es decir, prácticamente se ha duplicado la productividad de la planificación de inspecciones que resultan en algún tipo de infracción laboral.

INFRAESTRUCTURA

Con el apoyo del gobierno de Bielorrusia se desarrolló una aplicación móvil llamada [RoadLab](#), que detecta las necesidades de mejoramiento vial en base a una evaluación automática de la calidad de

⁶ El Laboratorio de Analítica Pública es una iniciativa conjunta del Centro de Investigación Avanzada en Educación de la Universidad de Chile (CIAE) y el Centro de Inteligencia Territorial (CIT) de la Universidad Adolfo Ibañez.

la superficie de las calles y caminos (World Bank Group, 2016). Para ello, se utiliza el acelerómetro de los teléfonos celulares que además entregan constantemente su posición usando el GPS. Posteriormente, un algoritmo divide las calles en segmentos de 100 metros, georreferenciando el inicio y el final de cada segmento, y luego envía la información en caso que se identifique algún hoyo en ese segmento. La aplicación diferencia entre un hoyo y otras contingencias de manejo (como una frenada brusca utilizando *ML*) (World Bank Group, 2016).

RECAUDACIÓN DE IMPUESTOS

Algunos organismos tributarios utilizan el análisis predictivo de sus propios registros para detectar probables casos de impuestos impagos o de prácticas contables engañosas (Goldsmith et al., 2016). En EE.UU., el estado de Nueva York utiliza el análisis predictivo para detectar devoluciones de impuestos cuestionables. Desde su implantación, el programa ha ayudado a que se recuperen 400 millones de dólares que se estaban perdiendo anualmente por devoluciones ilícitas, además de 100 millones de dólares derivados de una mayor disponibilidad de ingresos (Goldsmith et al., 2016). En Chile, a partir de agosto de 2017, el Servicio de Impuestos Internos ofrece una propuesta de declaración de IVA (a partir de distintas fuentes de información), que hoy es utilizada por alrededor de 700 mil contribuyentes. Esto facilita el cumplimiento, disminuyendo errores y dando certeza a los contribuyentes de cómo deben registrar sus operaciones y declarar sus impuestos. Esto ya produjo un aumento de recaudación de un 6,79% entre agosto 2017 y septiembre de 2017 y de un 6,41% comparado con septiembre de 2016. En cuanto a la cantidad de declaraciones, éstas acumulan hasta diciembre de 2017 un aumento de 15,94%. En la misma línea, el SII estima conductas de fraude relacionadas con el impuesto al valor agregado con uso de distintas técnicas de *data mining*, con tasas de acierto superiores al 80% (Castellón González & Velásquez, 2013).

RUTAS PARA VEHÍCULOS DE EMERGENCIAS

La Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile desarrolló una aplicación denominada *Commander* (Noticias Universidad de Chile, 2015) que permite gestionar el despacho de los carros del Cuerpo de Bomberos de Santiago por las rutas menos congestionadas. Esta información que se recopila a partir de los datos de más de 6 mil buses del Transantiago y el Cuerpo de Bomberos de Santiago cada 30 minutos. Con esta aplicación se ha logrado disminuir en hasta un 40% los tiempos de respuestas ante emergencias (Noticias Universidad de Chile, 2015).

SALUD PÚBLICA

- **Predicción de demanda de pacientes con VIH:** En Sudáfrica se generaron algoritmos que permiten unificar las bases de datos de distintas instituciones públicas que manejan información respecto a pacientes con VIH, tales como centros de salud que ofrecen asistencia en esta materia, y cantidad de pruebas de laboratorio realizadas para detectar esta enfermedad, entre otra información relacionada. De este modo, se puede identificar los lugares en que se atienden las mayores proporciones de pacientes con VIH a nivel nacional, provincial, distrital o de dependencia de salud, asignando de forma más eficiente los recursos e insumos necesarios para las atenciones de estos pacientes (World Bank Group, 2016).
- **Inspecciones sanitarias:** Un caso de estudio conjunto entre las Universidades de Stony Brook y Harvard (EE.UU) propone algoritmos de ML para mejorar la eficacia de las inspecciones de higiene asignando qué lugares visitar. Estos modelos predictivos se realizan en base al análisis de las reseñas en línea de restaurantes, donde se buscan las quejas asociadas a infracciones sanitarias (Kang, Kuznetsova, Luca, & Choi, 2013).

- **Plagas de ratones:** En Chicago, EE.UU., se desarrolló un piloto operacional a partir del análisis de datos del servicio telefónico 311⁷. Se detectó una relación entre ciertos tipos de llamadas y problemas de plagas de ratas. El Departamento de Calles y Saneamiento usó esta información para mejorar su estrategia para ubicar los sebos para las ratas en los lugares donde fuesen más efectivos (Goldsmith et al., 2016).
- **Reprogramación de cirugías riesgosas:** Una investigación conjunta de las Universidades de Cornell y Harvard mostró el potencial beneficio de usar algoritmos de ML para reasignar las operaciones de reemplazo de caderas y rodillas en la tercera edad (Kleinberg, Ludwig, Mullainathan, & Obermeyer, 2015). Priorizando las operaciones a aquellos pacientes con una mayor probabilidad de sobrevivir y menores complicaciones (Hamel, Toth, Legedza, & Rosen, 2008), se estimó que se podrían evitar 10.512 operaciones con posibilidad de futuras complicaciones y redistribuir 158 millones de dólares al año evitando operar a personas de mayor riesgo y evitando complicaciones asociadas a la cirugía (Kleinberg et al., 2015).

SEGURIDAD PÚBLICA

- **Crimen e infraestructura urbana:** En Bogotá, Colombia se estudió la asociación entre crimen e infraestructura urbana. Para esto se analizaron los datos de las rutas del sistema de buses '*Bus rapid transit*' (BRT), en conjunto con el modelamiento de terreno riesgoso ('*risk terra in modeling*' en inglés). Como resultado, se asociaron ciertas zonas cerca de hospitales, colegios, farmacias y estaciones de buses a asaltos y asesinatos, así como también se identificaron horas de mayor criminalidad, y se predijeron las áreas de la ciudad que sufrirán crímenes en el futuro (World Bank Group, 2016).
- **Delincuencia:** los departamentos de policía pueden usar analítica para predecir patrones delictivos, establecer planes de vigilancia, y estimar qué individuos o barrios estarán en riesgo (Goldsmith et al., 2016). En EE.UU existen numerosos ejemplos de uso de estos sistemas en ciudades como: Los Ángeles (delitos descendieron un 13% en las áreas donde se implementó un sistema similar), Chicago, Nueva York, Miami-Dade, Shreveport (ahorro de entre un 6%-10% en costos laborales y disminución de un 35% en delitos contra la propiedad), Charlotte-Mecklenburg, Memphis (reducción en más de un 30% de las delincuencia, incluyendo un 15% de delitos violentos en un periodo de 4 años), Arlington y Lancaster (Goldsmith et al., 2016). En este último caso se logró una reducción de la delincuencia en un 37% en 3 años y se logró un beneficio anual de la productividad de un millón de dólares (Goldsmith et al., 2016).

EN SINTESIS: La analítica de datos tiene aplicación en distintos ámbitos del sector público, posibilitando visualizar y accionar problemáticas que serían extremadamente costosas e incluso imposibles de abordar con métodos tradicionales. Producen impactos tanto en la relación costo-efectividad de la implementación de programas y políticas públicas, así como en la previsión de problemas en los que actualmente sólo se puede gestionar y mitigar sus consecuencias. Aunque varios de estos casos presentan desafíos éticos en la gestión y uso de datos públicos (aspectos de

⁷ El número 311 se utiliza en Estados Unidos para proveer acceso a servicios municipales no urgentes. En el caso de la ciudad de Chicago, contempla alrededor de 500 tipos de llamadas, desde quejas sobre restaurantes faltos de higiene, hasta baches en las calles (Goldsmith, Crawford, & Weinryb Grohsgal, 2016).

extrema relevancia pero que no forman parte de este estudio), el uso intensivo de datos en políticas públicas se presenta como una herramienta efectiva para una mejor gestión pública.

2 MARCO CONCEPTUAL

Como se puede observar a partir de los casos reseñados, existen muchas aplicaciones puntuales, útiles y muy rentables para generar valor público. La pregunta es, entonces: ¿cómo podemos escalar sustentablemente el desarrollo de la analítica de datos para la toma de decisiones y el desarrollo de políticas públicas en el sector público chileno?

Para ello, se requiere en primer lugar definir un marco conceptual que nos permita generar un diagnóstico del estado actual de *data-driven government* en Chile y a partir de las brechas detectadas, generar una estrategia con un *roadmap* claro que permita masificar el uso de la evidencia en la toma de decisiones y en el diseño, desarrollo, implementación y evaluación de las políticas públicas como una acción consistente y sistemática dentro del Gobierno de Chile.

En esta sección, desarrollaremos algunos conceptos claves para la discusión en las siguientes secciones: la cadena de valor de los datos, el ecosistema de generación de valor a partir de los datos, un modelo de madurez de gobierno basado en datos y finalmente un modelo de transformación de proyectos de modernización en el sector público. Estos modelos serán utilizados para proveer un marco teórico para el desarrollo de las siguientes secciones.

2.1 LA CADENA DE GENERACIÓN DE VALOR A PARTIR DE LOS DATOS

En primer lugar, analizaremos el concepto de **cadena de generación de valor a partir de los datos** (OECD, 2017a), es decir, como el dato producido o recolectado dentro del Estado se transforma en evidencia y **un activo estratégico** (sección 1.2) y qué acciones hay que desarrollar para que esto ocurra. La cadena de generación de valor a partir de los datos considera **8 pasos** que van desde la recolección y producción de los datos hasta su apertura como datos abiertos, re-utilización y ciclos de retroalimentación, como se muestra en la Figura 2.

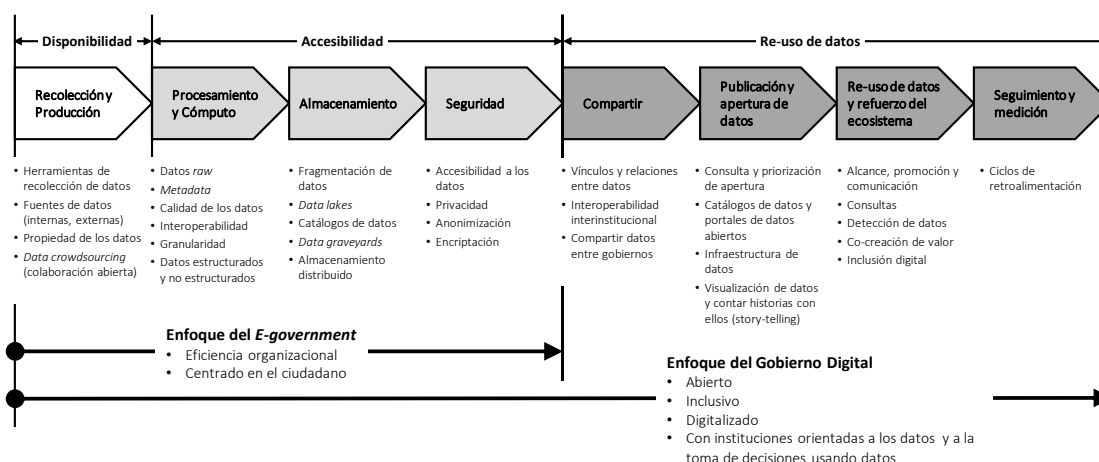


Figura 2: Modelo de cadena de generación de valor a partir de los datos. Adaptación de Tabla 5.1 creada por Rivera-Perez (OECD, 2017a).

La OCDE agrupa estos 8 pasos en 3 grandes dimensiones en el índice *Open-Useful Reusable Government Data Index* (OURdata): disponibilidad, accesibilidad y reutilización de datos (OECD, 2017b). Además, la cadena de generación de valor a partir de los datos nos permite diferenciar dos grandes enfoques sobre los procesos de digitalización del gobierno: el **gobierno electrónico**, que se focaliza principalmente en las dimensiones de disponibilidad y accesibilidad de datos; y el **gobierno digital**, que hace énfasis en el re-uso de datos producidos previamente por distintas agencias de gobierno para la toma de decisiones.

Para el sector público, es fundamental la gobernanza de la cadena de generación de valor a partir de los datos para transformar los datos en un **activo estratégico** (OECD, 2017a). Esto requiere entender cómo la cadena de generación de valor se vincula al logro de los objetivos generales de la política pública (OECD, 2017a). Por lo tanto, es vital concebir todas las etapas de la cadena de generación de valor y sus resultados (por ejemplo, registros de datos y datos abiertos de gobierno) como **elementos interconectados del proceso de producción de ese valor**, y reconociendo cómo algunos factores clave (por ejemplo, capital humano, tecnología y los modelos organizacionales) interactúan durante todo el proceso e influyen positiva o negativamente sobre éste (OECD, 2017a; Gonzalez-Zapata & Heeks, 2017).

El concepto de cadena de generación de valor a partir los datos permite levantar requerimientos mínimos que favorezcan desarrollar un sector público que utilice los datos como un insumo estratégico (OECD, 2017a). Esto significa asegurar la disponibilidad de un ambiente institucional, tecnológico, cultural y regulatorio que favorezca este proceso (OECD, 2017a; Gonzalez-Zapata & Heeks, 2017), de acuerdo a los principios de gobierno digital de la OCDE (OECD, 2014).

2.2 ECOSISTEMA DE GENERACIÓN DE VALOR A PARTIR DE LOS DATOS

El segundo concepto corresponde al **ecosistema de generación de valor a partir de los datos**, entendido como la forma en que se crea un circuito virtuoso de consumo de los datos producidos por el Estado. Este ecosistema es relevante porque nos permite entender cuál es el rol de otros posibles actores tanto en la producción de datos y generación de evidencia accionable para el gobierno, considerando que todavía no tenemos implementada una institucionalidad robusta de gobierno digital (sección 2.1).

En este modelo se identifican principalmente tres *stakeholders*: el sector público, el sector privado y la ciudadanía (Figura 3), centrado principalmente en el consumo de datos abiertos. El **sector público** produce datos que se publican en forma abierta para que los otros actores del ecosistema puedan utilizarlo: la ciudadanía y el sector privado. Incluso, el mismo sector público puede utilizar esos datos como forma de evitar los costos de transacción entre servicios, especialmente de tipo jurídico, para compartir información. De esta forma, los datos abiertos gubernamentales (DAG) funcionan como el “combustible” del sistema de valor que se crea con la participación de estos 3 actores, constituyendo un ecosistema de **innovación abierta** (Kankanhalli, Zuidervijk, & Tayi, 2017).

La **ciudadanía**—compuesta entre otras por organizaciones de la sociedad civil y no gubernamentales, la prensa y la academia— utiliza los datos abiertos con fines de producir evidencia sobre la transparencia, eficacia y eficiencia del accionar del gobierno, e investigación y desarrollo. Esto entrega retroalimentación al sector público sobre oportunidades de seguimiento y evaluación de políticas públicas para detectar, por ejemplo, ineficiencias en la implementación de programas, uso de los recursos, dolo, corrupción y colusión, asimetrías de información y falta de competencia en el

mercado, entre otros resultados. Además, la información de la propia ciudadanía puede generar valor para el Estado. Por ejemplo, iniciativas de datos colaborativos espaciales como *OpenStreetMap*⁸ permitirían al Estado mejorar los planos de las ciudades y la geolocalización de direcciones e incluso las mallas de transporte para producir modelos de tráfico.

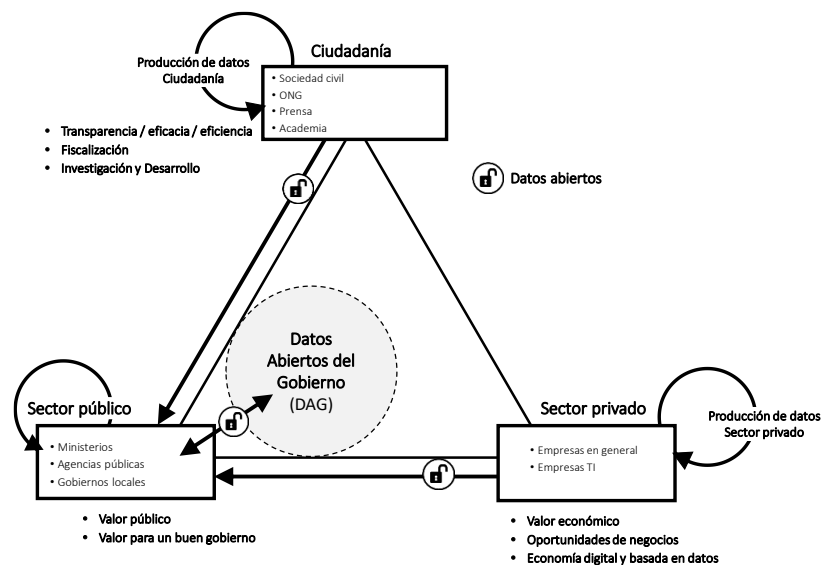


Figura 3: Modelo de generación de valor a partir de los datos: principales actores y sus roles. Adaptación y modificación propia de la Figura 5.1 del Capítulo 5 “Building a data-driven public sector in Norway”, OCDE (OECD, 2017a).

El **sector privado** busca oportunidades de generar valor económico a partir de datos públicos, detectar oportunidades de negocios y eventualmente el desarrollo de la economía digital y basada en datos. Por ejemplo, los datos abiertos del censo de población y vivienda en combinación con los de *OpenStreetMap* pueden usarse para detectar falta de diferentes servicios en territorios en expansión lo que constituye una oportunidad de negocio para cerrar dichas brechas. Actualmente son muchas las empresas y emprendimientos que están generando valor económico a partir de los datos públicos⁹, aunque a la fecha en Chile no se observan avances en esta línea [83] [86] [87].

Sin embargo, este ecosistema no funcionará si es que el estado no intenciona la producción del insumo que hace que este círculo virtuoso funcione. En este sentido, por ejemplo, cabe preguntarse cuáles son los criterios con que se publican los datos abiertos: ¿es sólo un trámite administrativo más? ¿son los datos publicados relevantes para aprovechar las capacidades que tienen los otros actores del ecosistema? ¿hay una intencionalidad de qué los actores del ecosistema respondan preguntas estratégicas? Esto significa que debe existir una estructura clara de gobernanza y de estrategia desde el sector público que permitan alimentar este ecosistema de producción de evidencia.

⁸ OpenStreetMap (<http://www.openstreetmap.org>) es “gran comunidad de colaboradores que con sus contribuciones al mapa añaden y mantienen datos sobre caminos, senderos, cafeterías, estaciones de ferrocarril y muchas cosas más a lo largo de todo el mundo” (OpenStreetMap, 2017).

⁹ Open Data’s Impact (<http://odimpact.org/>)

Para explicitar aún más las oportunidades para producir valor a partir de los datos, —no necesariamente abiertos y que requieren de un tratamiento especial en cuanto al aseguramiento de la protección de datos personales, con los correspondientes desafíos jurídicos que hay que abordar en la materia— hemos extendido el modelo del ecosistema tal como se muestra en la Figura 4.

En este caso, el **sector público** produce y consume datos a la vez utilizando el modelo de la cadena de generación de valor a partir de los datos (sección 2.1, Figura 2). Al incorporarse la cadena de generación de valor, se intenciona (en la dimensión de re-uso) la publicación y apertura de datos que incluye mecanismos de consulta a los actores del ecosistema y priorización de apertura de datos. La extensión del modelo de la Figura 4, permite entender además a los DAG como un mecanismo que permite hacer funcionar el ecosistema de generación de valor a partir de los datos, dándole una funcionalidad estratégica más allá de la (necesaria) transparencia de la función pública.

Por lo tanto, además de existir una explícita gobernanza de la cadena y ecosistema de generación de valor, se debe generar incentivos para el trabajo colaborativo del sector privado junto con las universidades y centros de investigación como una forma efectiva de abordar problemas complejos de interés público (Tomar, Guicheney, Kyarisiima, & Zimani, 2016). Además, el sector privado y la academia pueden ser una buena fuente de conocimientos, capacidades técnicas e incluso de datos para resolver desafíos de política pública (Tomar et al., 2016), tal como se ilustró en distintos ejemplos en la sección 1.4. Este diálogo y sinergia permiten que se fortalezcan las expectativas de largo plazo y esta política trascienda los cambios de gobierno (Corduneanu-Huci, Hamilton, & Ferrer, 2012).

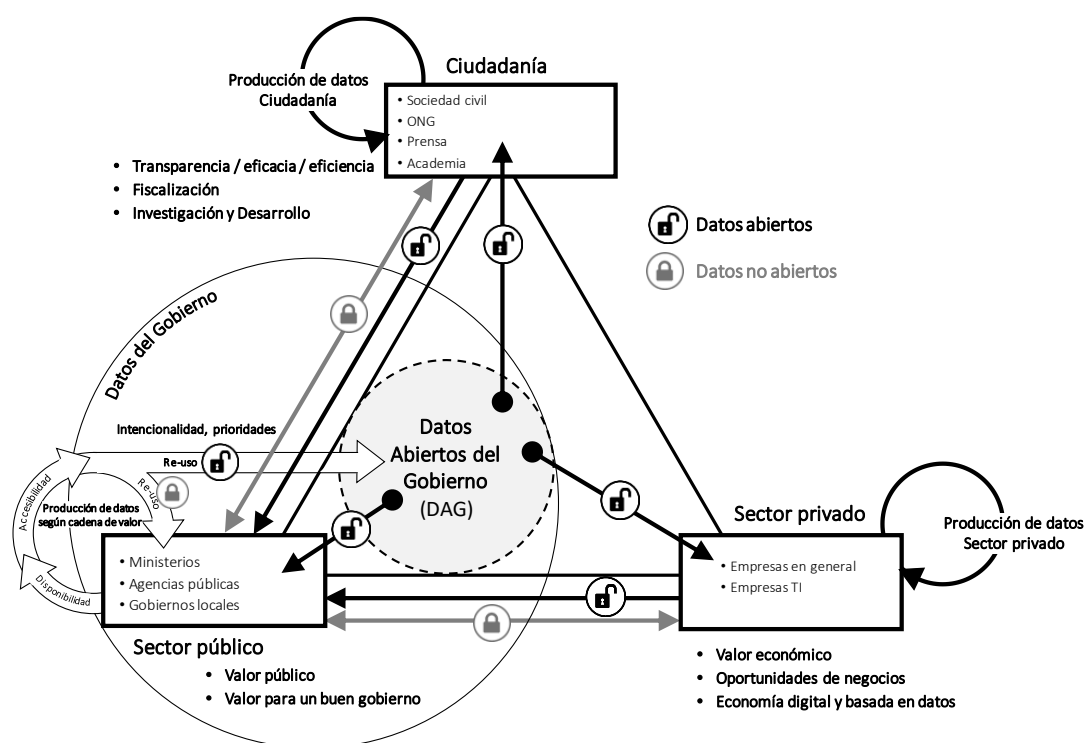


Figura 4: Modelo extendido de generación de valor a partir de los datos: principales actores y sus roles. Adaptación y modificación propia de la Figura 5.1 del Capítulo 5 "Building a data-driven public sector in Norway", OCDE (OECD, 2017a).

2.3 MODELO DE MADUREZ DE GOBIERNO BASADO EN DATOS

Desarrollar la cadena de generación de valor, promoviendo la creación de un ecosistema de generación y consumo de datos es, sin duda, un gran desafío. Sin embargo, nuestro país no está comenzando desde cero, ya que cuenta con un portal de datos abiertos (datos.gob.cl) que depende de la División de Gobierno Digital del Ministerio Secretaría General de la Presidencia (SEGPRES). Plataforma donde distintas reparticiones públicas publican sus datos, y que además permite alimentar algunas aplicaciones desarrolladas por las instituciones públicas (app.gob.cl).

Por lo tanto, es necesario levantar un diagnóstico de cuentas sobre en qué lugar nos encontramos, que puede variar dependiendo de la repartición pública analizada. Wiseman (2016) propone un modelo de madurez que permite vincular el desarrollo de la cadena de generación de valor para desarrollar tanto el concepto de gobierno digital como el ecosistema de generación de valor a partir de los datos. Este modelo de madurez consiste en una rúbrica de 5 etapas, que se muestran en la **Tabla 2**.

Tabla 2: Modelo de madurez de gobierno basado en datos (Wiseman, 2016).

Etapas de madurez de gobierno basado en datos			
1. Publicar	2. Refinar	3. Analizar	4. Optimizar
<ul style="list-style-type: none"> • Inicialmente la plataforma de datos abiertos de gobierno (DAG) incluye datos que se pueden publicar con poco esfuerzo: <ul style="list-style-type: none"> ○ Públicos por reglamento. ○ Frecuentemente pedidos por transparencia. ○ Que ya estén en formato legible por máquinas o en formato estándar. • Usuarios descargan los datos y hacen sus propios análisis y visualizaciones. • Se definen estrategias y políticas de datos abiertos. • Instituciones públicas tienen gran diversidad en disponibilidad y calidad de los datos. • Datos para gestión del desempeño están disponibles en algunas instituciones, pero no es uniforme y no son publicados típicamente. 	<ul style="list-style-type: none"> • DAG es periódicamente actualizada y mejorada. • Se asegura la calidad y consistencia a través de auditorías u otros métodos de retroalimentación. • Plataforma de DAG es estable y soporta grandes volúmenes de datos. • Transparencia se logra incrementalmente. • Existen canales formales que permiten entregar retroalimentación para la mejora en la disponibilidad de datos. • DAG permite mayor y mejor compromiso cívico. • Existe personal capacitado a nivel centralizado y en otras reparticiones públicas. • Se incorpora datos de entidades semi-públicas. • Disponibilidad de APIs para uso de aplicaciones externas • Datos para la gestión del desempeño aumentan su disponibilidad y se comparten públicamente. 	<ul style="list-style-type: none"> • DAG está bien organizada y es fácil de navegar, descargar y usar para investigación y análisis de la ciudadanía. • DAG provee herramientas de análisis y visualización. • Existe un encargado de analítica con una clara misión y recursos específicos. • Se usa analítica en algunas reparticiones por parte de algunos líderes. • Algunos proyectos intersectoriales de analítica. • Éxitos documentados generan interés. • Crece una cultura de datos entre distintas reparticiones. • Crece el desarrollo de habilidades de análisis de datos dentro del gobierno y las oportunidades de capacitación. • DAG y datos para programas estadísticos, gestión del desempeño son de alta calidad y consistentes entre reparticiones, permitiendo análisis intersectoriales robustos. • Resultados de la gestión operacional y del desempeño están bien integrados a las actividades de analítica de datos. 	<ul style="list-style-type: none"> • Las decisiones clave de política pública son tomadas en base a la evidencia provista por el análisis de datos • La analítica de datos permita la asignación de recursos a las necesidades públicas más altas. • Los liderazgos de las reparticiones confían en la analítica de datos para la toma de decisiones clave, con una mentalidad del “problema primero” y buscando datos para poder responder preguntas. • Habilidades de analítica de datos y científicos de datos existen en el gobierno central y otras reparticiones. • Reparticiones colaboran en proyectos intersectoriales de analítica de datos. • Proliferan casos de éxito propios o de terceros, y algunos incluyen lineamiento para replicarlos. • Se optimiza la gestión del desempeño, con un fuerte compromiso de la ciudadanía para entregar retroalimentación al gobierno, lo que mejora de la transparencia y confianza en el gobierno.

La etapa inicial (*Publicar*) se refiere a la existencia de los datos abiertos y su gestión centralizada, luego este modelo va complejizando tanto la forma en cómo se generan los datos abiertos en cuanto a consistencia y calidad, así como las capacidades al interior del Estado para incorporar la analítica como actividad sistemática de la función pública (*Optimizar*). Este modelo será utilizado para diagnosticar la situación de nuestro país en relación con el estado de implementación del gobierno basado en datos.

2.4 MODELO DE TRANSFORMACIÓN DIGITAL EN EL SECTOR PÚBLICO

El último modelo que proponemos antes de realizar un diagnóstico y plantear una estrategia para el desarrollo de un gobierno digital guiado por datos (según la definición de la cadena de generación de valor en sección 2.1) es un modelo de transformación digital en el sector público. Este modelo, que ha sido desarrollado por el Programa de Modernización del Sector Público del Ministerio de Hacienda (Ledermann & Madariaga, 2017), conceptualiza las dimensiones esenciales que requiere cualquier proyecto de transformación digital para ser exitoso: *liderazgo político*, *gobernanza y capacidad institucional*, y *presupuesto* (Figura 5). La intersección de estas tres dimensiones permite la transformación institucional, dan sustentabilidad al proyecto, y finalmente movilizan la política pública para que sea exitosamente implementada.

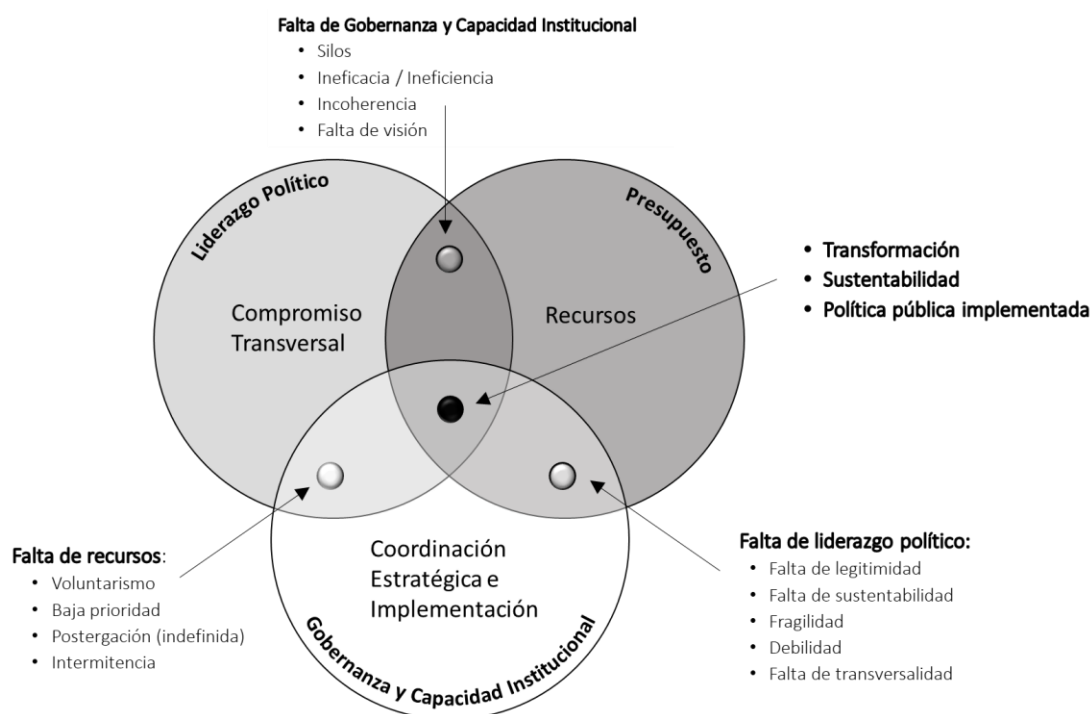


Figura 5: Modelo de transformación digital en el sector público (Ledermann & Madariaga, 2017).

El **liderazgo político** aporta un compromiso transversal y da legitimidad al proyecto desde la autoridad del servicio ante los diferentes actores políticos involucrados. Cuando este liderazgo existe, la institución se moviliza alineadamente para la implementación de la iniciativa de gobierno digital. Sin embargo, si este liderazgo no existe, es posible cuestionar la legitimidad del proyecto, que lo hace

dependiente de las voluntades de otros actores que no necesariamente tienen un alto nivel de compromiso. En este último contexto, las iniciativas se vuelven frágiles y vulnerables ante la contingencia de la operación, aumentando, entre otros, los costos de transacción para su implementación.

La **gobernanza y capacidad institucional** generan la coordinación estratégica y operacional de la institución para que la transformación digital ocurra. Esta gobernanza contribuye a la sincronización adecuada de las iniciativas en el marco de complejos y muchas veces lentos y rígidos procesos burocráticos. Cuando estos aspectos no están presentes, existe el riesgo de que el proyecto se implemente en forma compartimentada, sin coherencia (por falta de comunicación), de forma ineficaz e ineficiente (porque no están todos los actores involucrados), lo que en definitiva puede causar la falta de visión sistémica y capacidades coordinadas que pueden amenazar la implementación exitosa del proyecto.

Finalmente, la gestión estratégica del **presupuesto** disponibiliza los recursos necesarios para la implementación del proyecto de transformación digital. Cuando estos recursos existen, las instituciones pueden planificar de forma estratégica sus procesos de transformación digital. Sin embargo, cuando no existen dichos recursos, el proyecto tiene una baja prioridad porque compete con las tareas de la contingencia y las lógicas inerciales de financiamiento público, lo que produce intermitencia en la implementación. Esto hace que el proyecto avance simplemente por el voluntarismo de los actores involucrados o caiga en la postergación propia de la baja prioridad o simplemente no comience nunca.

3 DIAGNÓSTICO

Esta sección tiene por objetivo analizar el estado actual del uso de datos para políticas públicas en Chile, y en particular identificar las brechas que impiden la implementación de una política efectiva de *data-driven government* en el país. Este ejercicio es realizado utilizando como referencia el *roadmap* de gobierno basado en datos desarrollado por Wiseman (2016) (Tabla 2 en sección 2.3), y que define cuatro etapas para la implementación de una adecuada política de uso intensivo de datos en políticas públicas: *Publicar*, *Refinar*, *Analizar*, y *Optimizar*. Es importante destacar que la autora sugiere este *roadmap* para definir una estrategia de desarrollo de políticas públicas basadas en datos, pero no representa la progresión secuencial de éstas. Por el contrario, es posible identificar logros y desafíos en cada una de ellas, lo que da cuenta de la multidimensionalidad de estas iniciativas, y su desarrollo muchas veces inorgánico al interior de las instituciones públicas. Además, se han considerado diversos indicadores de datos públicos, los que se concentran principalmente en *open data*. Sin embargo, se identifica la falta de suficientes estudios y/o diagnósticos sobre el uso avanzado de datos en las instituciones públicas chilenas, aspecto que debiera ser incorporado como parte de futuros estudios para la implementación de la estrategia propuesta en la sección 4.

3.1 PUBLICAR: DISPONIBILIZAR DATOS PARA LA CIUDADANÍA Y LAS INSTITUCIONES PÚBLICAS

En esta primera etapa hacia un sistema de políticas públicas basadas en datos, el gobierno central y/o las instituciones públicas definen políticas y estrategias de datos abiertos, donde se prioriza la apertura de datos de alta demanda por la ciudadanía o que se encuentran fácilmente disponibles en

formatos legibles por máquina. Estos *datasets* de rápida disponibilización ayudan a generar apoyo político y minimizan los esfuerzos iniciales de las instituciones públicas para su publicación, al mismo tiempo que pueden responder a las necesidades principales de quienes más frecuentemente consumen estos datos. Sin embargo, en esta primera etapa de madurez las instituciones públicas pueden tener diferentes niveles de calidad y disponibilidad de datos debido a la falta de políticas centrales de gobernanza de datos y los marcos normativos vigentes, donde por lo tanto predominan esfuerzos sectoriales con diferentes niveles de calidad y pertinencia.

En el caso de Chile, a la fecha no existe una política clara y definida de datos abiertos en el gobierno central. Aunque la División de Gobierno Digital de SEGPRES (DGD, anteriormente conocida como Unidad de Modernización y Gobierno Digital) ha hecho esfuerzos en la materia durante los últimos cinco años, todavía no existe una política que defina roles, compromisos, presupuesto y responsabilidades tanto para la propia División como para las instituciones que disponibilizan datos (González-Zapata & Heeks, 2017; OECD, 2016). La falta de una política consistente de datos abiertos se manifiesta en el irregular desempeño de la iniciativa; teniendo un portal de datos desde el año 2012 (datos.gob.cl), la publicación de *datasets* ocurre de forma irregular, con bajos niveles de calidad e integridad de datos, y de acuerdo al énfasis político otorgado por cada institución pública. A la fecha, el único documento con capacidad regulatoria en la materia es el Instructivo Presidencial 05 (Gobierno de Chile, 2012), que solicitó la publicación de cinco *datasets* en un período de 60 días en el año 2012, pero que no determina las frecuencias de publicación, ni los formatos y licencias asignadas a estos conjuntos de datos. La División desarrolló una guía de publicación de datos (Ministerio Secretaría General de la Presidencia, 2013a) y una norma técnica de datos abiertos (Ministerio Secretaría General de la Presidencia, 2013b), pero que actúan únicamente como recomendaciones técnicas y no tienen capacidad normativa. Como resultado, la publicación de *datasets* por parte de las instituciones públicas ocurre de forma irregular, y con heterogéneos niveles de calidad y relevancia, lo que puede resultar en iniciativas basadas en datos de bajo uso e impacto (González-Zapata, 2016).

En la ausencia de una política central de apertura de información, varias instituciones públicas han decidido desarrollar sus propios proyectos de datos abiertos para tener mejor control de los datos que son publicados, así como orientar la interpretación y uso de éstos. Estos portales frecuentemente concentran tanto la disponibilidad de *datasets* en formatos abiertos, así como visualizaciones específicas según el contexto de negocio de cada institución, y proveen de mejores métodos de interpretación para estos datos versus una plataforma centralizada. Este es el caso de proyectos como Energía Abierta¹⁰, Junaeb Abierta¹¹, Minería Abierta¹² o el portal de datos de ChileCompra¹³, entre otros. Aunque son proyectos de gran valor público, representan un número menor de instituciones públicas con políticas de apertura de información pública (The World Wide Web Foundation, 2016).

El marco regulatorio de acceso a la información en Chile asegura que todos los actos administrativos de las instituciones del Estado y sus antecedentes son públicos, y por ende deben estar disponibles para los ciudadanos. Este principio constitucional (Ministerio Secretaría General de la Presidencia,

¹⁰ <http://energiaabierta.cl/>

¹¹ <http://junaebabierta.junaeb.cl/>

¹² <http://www.mineriaabierta.cl/>

¹³ <http://datosabiertos.chilecompra.cl/>

2005) está operacionalizado por la ley de transparencia 20.285 que define tanto la publicidad de la información como los métodos de acceso a ésta (pasiva o activa) (Ministerio Secretaría General de la Presidencia, 2008). Siendo la ley una herramienta para determinar la publicidad de los datos públicos, principios esenciales de datos abiertos como la apertura por defecto o por diseño no están contemplados en la ley, así como tampoco el formato y tipos de licencia sobre los cuales la información pública debe ser entregada. Como resultado, las instituciones públicas han adoptado la práctica de responder mayoritariamente a las solicitudes de transparencia en formatos no legibles por máquina – tanto por su naturaleza jurídica como por la conveniencia de resguardar la interpretación de la información contenida en estos documentos, lo que dificulta la publicación bajo los principios de datos abiertos.

El resultado de este limitado marco institucional para la implementación de una política de datos abiertos es que Chile muestra niveles insuficientes de apertura de datos para su uso inteligente en la formulación y optimización de políticas públicas. Tal como indica el *OURdata Index* de la OCDE, índice que mide el nivel de desarrollo de las iniciativas de datos abiertos en sus países miembro, Chile es uno de los países con iniciativas de datos abiertos de menor desarrollo e impacto (OECD, 2017b), en particular dada la baja disponibilidad de datos de calidad y de políticas que fomenten su uso inteligente (ver Figura 6). Adicionalmente, otras mediciones internacionales de datos abiertos tales como el *Open Data Barometer* (The World Wide Web Foundation, 2017) o el *Global Open Data Index* (Open Knowledge International, 2017) sitúan a Chile entre los países con disponibilidad media de *datasets*, pero sin políticas e iniciativas de gobierno tendientes a asegurar su calidad e integridad, lo que produce como resultado bajos niveles de impacto en diferentes áreas definidas en estas mediciones. Como resultado, en el contexto de un análisis comparado con países que han llevado a cabo significativos esfuerzos para implementar políticas de *data-driven government*, Chile se encuentra aún en un nivel de madurez insuficiente de apertura de información para alcanzar un sistema robusto de políticas públicas basadas en evidencia.

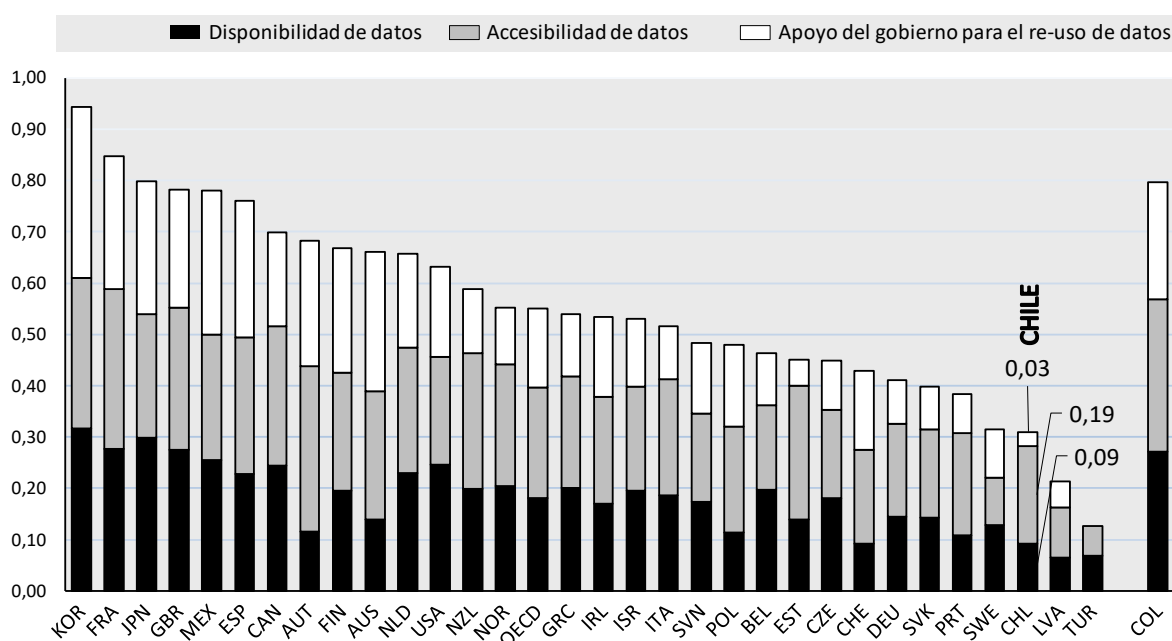


Figura 6: Chile en el Índice OCDE de datos públicos abiertos-útiles-reutilizables (OURdata) 2017 (OECD, 2017a, 2017b).

3.2 REFINAR: ASEGURAR LA CALIDAD DE LOS DATOS PÚBLICOS

En esta segunda etapa de madurez, las iniciativas de datos abiertos coexisten con políticas definidas y centralizadas de gobernanza de datos que permiten asegurar la calidad y relevancia de los datos que gestionan las instituciones públicas, y que por ende tienen un impacto en la madurez de la apertura de datos y su análisis inteligente. Este nivel además considera el mejoramiento continuo de la calidad de datos a través de auditorías internas, así como retroalimentación de la sociedad civil y otros *stakeholders*, lo que conlleva establecer métodos específicos para que estos *datasets* sean regularmente actualizados y mejorados.

En el contexto de Chile, el gobierno central carece de una política consistente y clara de gobernanza de datos, y las responsabilidades asociadas a la correcta gestión de los datos públicos se encuentran (si existen) distribuidas en diferentes instituciones públicas (ver Figura 7). Por ejemplo, respecto de las prácticas de captura, almacenamiento y explotación de datos, cada institución pública trata sus datos de acuerdo a los marcos regulatorios sectoriales, que en algunos casos están estipulados en sus respectivas leyes orgánicas —por ejemplo, Instituto Nacional de Estadísticas (INE), Registro Civil, y Ministerio de Desarrollo Social (MDS), entre otros— las que frecuentemente regulan la captura de datos desde otros servicios, pero no la disponibilización de sus propios *datasets*. Además, mucha de la información que las instituciones manejan está construida en papel y sin principios de gestión documental electrónica (Microsystem, 2016). Debido a la sensibilidad y relevancia de los datos que utilizan, cada institución ha históricamente definido políticas y principios que les permiten operar sus datos a nivel sectorial, pero que fomenta la lógica de silos de gestión de datos. Esta política inorgánica frecuentemente impide que los datos públicos puedan interoperar de forma eficiente.

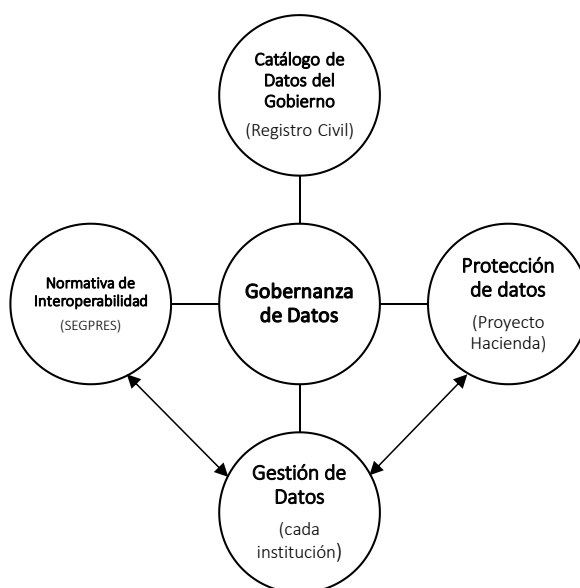


Figura 7: Esquema de gobernanza de datos en Chile

En términos de políticas de interoperabilidad, SEGPRES es actualmente la institución responsable de la implementación y fomento del uso de la Plataforma de Interoperabilidad del Estado (PISEE). Este sistema tiene por objetivo proveer de la infraestructura técnica y de los metadatos suficientes para

que las instituciones públicas puedan intercambiar información relevante para provisión de servicios públicos. Sin embargo, actualmente la iniciativa de interoperabilidad no es obligatoria para las instituciones públicas, y por lo tanto solo está siendo implementada por quienes voluntariamente adhieren a la normativa, reduciendo la posibilidad de explotar el cruce de datos por métodos más expeditos que la firma de convenios de colaboración o solicitudes directas entre jefes de servicio. Desde el 2014, la interoperabilidad se encuentra regulada por el Decreto del Ministerio de Economía Nº14 (2014), el que define los protocolos y métodos de intercambio de documentos electrónicos. Durante el año 2017 la División de Gobierno Digital trabajó para aprobar una nueva norma técnica para la interoperabilidad entre las instituciones del Estado, la que a la fecha de este documento todavía no ha sido publicada (Ministerio Secretaría General de la Presidencia, 2017).

Otro aspecto relevante para el adecuado intercambio de información entre los servicios del Estado es la protección de la vida privada y los datos personales. Chile actualmente cuenta con la ley 19.628 que regula la manipulación de datos por parte de las instituciones públicas y privadas, pero que dada su naturaleza de control sobre documentación tradicional (en papel) no contempla aspectos relevantes para el intercambio de información electrónica (Ministerio Secretaría General de la Presidencia, 1999). Además, la ley no se adecua a los estándares internacionales de seguridad de los datos personales, y deja brechas de interpretación que muchas veces actúan como barrera legal para la apertura y uso de datos gubernamentales en el desarrollo de políticas públicas. En la búsqueda de un marco regulatorio idóneo a las necesidades actuales, actualmente una nueva ley de protección de datos personales está siendo discutida en el Congreso, y que se espera pueda ser aprobada durante el primer semestre del año 2018.

Finalmente, respecto al registro de los tipos de bases de datos con datos personales disponibles en el Estado, es por ley el Servicio de Registro Civil e Identificación quien está mandado por ley para tener un catastro de todas las bases de datos personales de las instituciones públicas (Ministerio Secretaría General de la Presidencia, 1999). Sin embargo, el servicio no ha cumplido con este requerimiento legal a la fecha, y por ende se desconoce el número, tipo y alcance de las bases de datos del Estado, lo que evidentemente impide evitar la duplicidad en la recolección de datos, la priorización de datos para su apertura, y la cobertura de necesidades de datos por parte de las instituciones públicas.

Es en este limitado contexto es que el Estado chileno no tiene políticas de gobernanza de datos centralizadas y que actúen de forma efectiva como un marco normativo y operacional para la adecuada apertura y uso de datos públicos. Esta brecha es también constatada por el reciente informe del CEP sobre la modernización del sector público, y plantea la necesidad de definir un *Chief Data Officer* (CDO) que tenga capacidades regulatorias para dar cumplimiento y actualizar la normativa vigente y para que introduzca las mejores prácticas internacionales en gestión de datos en el Estado (Centro de Estudios Públicos, 2017). Este estudio sugiere estrategias en esta dirección en la sección siguiente.

Como parte de una adecuada estrategia de gobernanza de datos, los datos abiertos disponibles a la sociedad civil y las instituciones públicas debiesen estar frecuentemente bajo escrutinio de los usuarios que explotan estos datos. Estos procesos de mejoramiento continuo de datos públicos requieren establecer mecanismos donde sus usuarios frecuentes puedan reportar errores, mejoras, y nuevos *datasets* a ser abiertos. Bajo este modelo de mantenimiento de los datos como infraestructura pública, tanto usuarios de la sociedad civil como de las instituciones públicas pueden contar con datos de alta calidad y de acuerdo con sus requerimientos de uso. Este es el caso de ChileCompra, institución que encargó un estudio para evaluar la calidad de sus *datasets* y estableció un plan de mejora que permitan mantener altos estándares de calidad y disponibilidad de datos en

el futuro, en particular dada la pronta adopción del estándar internacional *Open Contracting Data Standard* (Development Gateway, 2017).

El resultado de este limitado marco institucional para la gobernanza de datos es que muchas veces actúa como una brecha entre las necesidades de información y su disponibilización. La dispersión de reglamentación y responsabilidades sobre los datos frecuentemente tiene un efecto atemorizante entre los tomadores de decisiones, quienes advierten la complejidad política de abrir y reutilizar datos sobre los cuales no hay principios de integridad ni calidad, así como de los costos de establecer mecanismos de interoperabilidad efectiva con otras instituciones. Como resultado, el actual modelo de gobernanza de datos opera como barrera para la apertura y uso inteligente de datos por parte de las instituciones públicas, y requiere de significativos esfuerzos en presupuesto, liderazgo y coordinación política para adoptar mejores estándares que faciliten estos procesos.

3.3 ANALIZAR: EXPLOTAR EL CONOCIMIENTO CONTENIDO EN LOS DATOS PÚBLICOS

El tercer nivel del modelo de madurez contempla estrategias para el uso inteligente de datos para políticas públicas. En particular, las instituciones debiesen contar con políticas internas y sectoriales de análisis de datos, así como con funcionarios públicos con las habilidades técnicas necesarias para hacer uso inteligente de estos datos a través de técnicas avanzadas de análisis como *Machine Learning* u otros tipos de inteligencia artificial. Además, en esta etapa los gobiernos tienen iniciativas maduras de apertura de información, las que retroalimentan el sistema de políticas públicas basadas en evidencia: dada la alta disponibilidad de datos públicos relevantes, el costo de transacción para aquellos que quieran hacer análisis avanzado de datos es bajo dado que los *datasets* tienen niveles suficientes de calidad e integridad, y se encuentran disponibles para todo uso bajo licencias abiertas, siempre respetando la regulación vigente y en particular la privacidad de las personas. Aunque aún a nivel sectorial, este estadio de madurez refleja la adopción cultural del análisis de datos como parte esencial del desarrollo de las políticas públicas de cada institución.

A la fecha, Chile se encuentra lejos de alcanzar este nivel de madurez. Según el *OURdata Index* de la OCDE, destaca la falta de políticas por parte del sector público chileno para fomentar el re-uso de la información, no sólo por parte de las instituciones públicas, sino también por la sociedad civil y la academia. La inexistencia de un adecuado ecosistema de reutilización de datos se explica por la ausencia de alianzas e incentivos (no necesariamente económicos, sino por ejemplo respecto a la facilitación y efectividad de la prestación de servicios o de las tareas de los funcionarios públicos a través de integración y análisis de datos) para que exista una adecuada reutilización de los datos públicos.

Respecto al uso inteligente de datos, en la actualidad responde a incentivos sectoriales y a la visión de tomadores de decisiones que en particular advierten las ventajas estratégicas de explotar el valor contenido en sus datos. Algunas instituciones públicas, apoyadas por el Programa de Modernización del Sector Público, han comenzado a implementar políticas de este tipo, tales como la Dirección del Trabajo (aumento de la tasa de éxito de las fiscalizaciones), la Superintendencia de Seguridad Social (predicción de reclamos en base a expedientes electrónicos), el Servicio Nacional del Consumidor (predicción de los casos de reclamo), o el Observatorio de Compras Públicas de ChileCompra (monitoreo inteligente y detección de irregularidades en los procesos de compra pública). En estos casos, las instituciones públicas han debido sortear diversas dificultades para realizar este nivel de análisis, desde que no contaban con la infraestructura digital y el capital humano necesario, hasta la integración con otros servicios para poder refinar sus modelos de análisis para lo cual deben firmar

acuerdos formales de colaboración caso a caso. Una de las principales dificultades observadas es la falta de un repositorio centralizado de datos del Estado, que permitiría fácilmente acceder a registros en tiempo real para alimentar los modelos analíticos de estas instituciones. En línea con este desafío y con las principales tendencias de gestión de datos públicos a nivel mundial, el CEP plantea la necesidad de gestionar los datos bajo el concepto de infraestructura pública, y desarrollar una plataforma integrada de datos (Centro de Estudios Públicos, 2017).

Otra causa del bajo nivel de adopción del análisis avanzado de datos es la falta de conocimiento de los tomadores de decisiones para identificar las ventajas estratégicas del análisis avanzado de datos públicos, y por ende estas instituciones carecen de recursos económicos y técnicos para llevar adelante proyectos de este tipo. Una estrategia importante para cerrar esta brecha es continuar los esfuerzos por difundir y capacitar a los funcionarios públicos.

En la ausencia de capacidades en los funcionarios públicos, datos de calidad, y del *awareness* entre los tomadores de decisiones de la potencialidad del análisis avanzado de datos, la sociedad civil y los centros de investigación han ayudado a cerrar esta brecha al actuar como intermediarios en la cadena de generación de valor a partir de los datos. Esto ayuda a dar sentido a los datos para que otros puedan entender el valor contenido en éstos, y obtener conclusiones que puedan informar a los tomadores de decisiones. Estos actores muchas veces disponibilizan los recursos y competencias ausentes en las instituciones para crear valor público a partir de conjuntos de datos que frecuentemente no conversan entre sí. En Chile, algunos casos que podemos mencionar son los de la Fundación Ciudadano Inteligente, el Observatorio del Gasto Fiscal y el Laboratorio de Analítica Pública, entre otros. Queda de manifiesto el rol que han jugado a la fecha distintas comunidades de la sociedad civil en Chile en reutilizar la información pública. Sin embargo, ante la falta de políticas de reutilización de datos, no existe a la fecha un ecosistema articulado de agentes re-utilizadores de datos públicos tanto dentro como fuera del Estado que inyecten las capacidades necesarias para hacer análisis intensivo de estos datos públicos. Es tarea del sector público el establecer alianzas con estas instituciones para explotar los datos que generan, a la vez que puedan retroalimentar sus modelos de análisis y beneficiarse de los resultados para la mejor toma de decisiones.

3.4 OPTIMIZAR: ANÁLISIS AVANZADO DE DATOS COMO POLÍTICA DE ESTADO

En este nivel de madurez, el análisis avanzado de datos es parte del *core* de negocio de las instituciones públicas, y es considerado como una política transversal, con liderazgo centralizado y con recursos suficientes para establecer proyectos intersectoriales. Pero aún más importante, *las decisiones clave de política pública son tomadas en base a la evidencia provista por el análisis de datos* (ver Tabla 2). En base a iniciativas de datos abiertos robustas y centradas en el usuario, a las capacidades técnicas y los recursos para hacer analítica, y al liderazgo y coordinación política para mover las decisiones clave desde la intuición política a la evidencia, esta etapa representa alcanzar el nivel de *data-driven government*.

Como la evidencia de las etapas anteriores muestra, Chile se encuentra aún lejos para alcanzar este nivel de madurez. La iniciativa de datos abiertos no es lo suficientemente robusta, no existen aún las capacidades y los recursos suficientes para llevar a cabo proyectos de análisis de alta envergadura, y principalmente aún no hay suficiente interés político para movilizar estos recursos y generar las coordinaciones intersectoriales necesarias. Es objetivo de este trabajo generar el interés de los tomadores de decisiones públicos para adoptar estas políticas, y por ende movilizar los esfuerzos hacia un *sistema de análisis de datos públicos*. La siguiente sección justamente tiene como propósito

el diseñar un sistema de datos para el Estado chileno, que responda a las limitaciones institucionales existentes, y que genere tanto una estrategia de largo plazo para alcanzar el *data-driven government*, así como para generar incentivos a través del desarrollo de iniciativas de corto plazo que muestren los beneficios de adoptar estas prácticas.

3.5 DIAGNÓSTICO A TRAVÉS DEL MODELO DE TRANSFORMACIÓN DIGITAL

La evidencia presentada a lo largo de esta sección sugiere que el sector público chileno aún se encuentra a una distancia significativa del estado de transformación digital en el uso intensivo de datos para las políticas públicas. De acuerdo con el modelo de transformación digital, el uso intensivo de datos puede ser analizado según los niveles de liderazgo, gobernanza y capacidad institucional, y presupuesto, los que en su conjunto determinan el estado ideal de transformación para que esta política ocurra. Nuestro análisis de esta situación se grafica en la Figura 8.

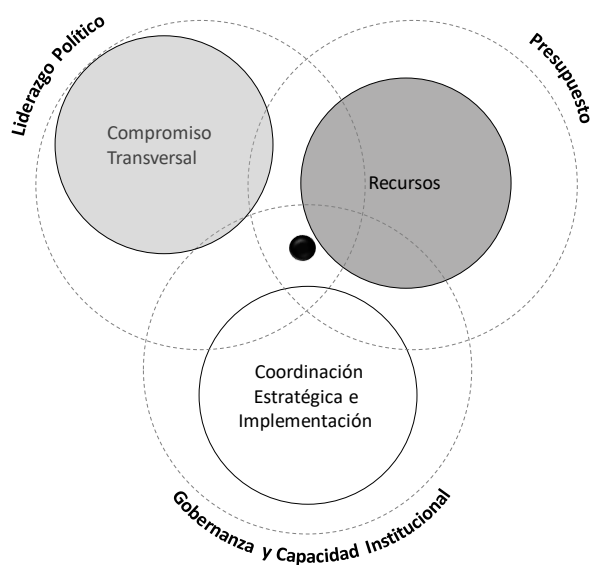


Figura 8: Diagnóstico del uso intensivo de datos bajo el modelo de transformación digital

Respecto al **liderazgo político**, el énfasis otorgado por las autoridades de diferentes administraciones no ha sido suficiente para relevar la importancia de implementar una estrategia de apertura y uso de datos transversal al Estado de Chile, así como para que los funcionarios públicos entiendan el valor de estos datos tanto para la propia institución como para la gestión de todo el Estado. Aunque es posible observar que lentamente las autoridades advierten los beneficios del uso intensivo de datos, se requiere mayor liderazgo y convencimiento político para desarrollar una estrategia que permita su implementación efectiva en las diferentes instituciones públicas. El rol coordinador de la División de Gobierno Digital debiera ser potenciado de forma que sea habilitante para definir e implementar políticas y normas técnicas tanto para la apertura como para el uso de datos. Como resultado, a la fecha los esfuerzos en apertura y uso de datos se concentran en iniciativas sectoriales, con diferentes niveles de calidad y relevancia.

Por otra parte, en la dimensión de **gobernanza y capacidad institucional**, las diferentes iniciativas de analítica de datos han sido predominantemente sectoriales, lo cual implica que todavía el sector público en este ámbito está organizado en silos y de forma inorgánica. Falta una política centralizada y con autoridad de seguimiento para las diferentes iniciativas que surjan y que permita acompañar, aprender de ellas y hacerlas coherentes. También existe una alta heterogeneidad en cuanto a las capacidades y las estructuras institucionales de apertura y uso de datos, lo que producirá resultados disímiles al impulsar este tipo de proyectos.

Finalmente, en la dimensión de **recursos**, éstos han estado focalizados en iniciativas específicas y sin una visión de uso sistemático de la analítica como función del Estado e instrumento de modernización a nivel general. Si bien se han disponibilizado recursos para la apertura de datos y para experiencias de uso intensivo a nivel sectorial, son aún insuficientes para implementar una agenda transformadora de uso intensivo de datos en el sector público.

Como se ilustra en la Figura 8, ninguna de las dimensiones tiene la importancia suficiente para movilizar la política de apertura y uso intensivo de datos en la dirección esperada, lo que significa que para que ésta pueda ser implementada se requieren de esfuerzos en las tres direcciones, aunque prioritariamente se requieren mayores niveles de liderazgo político y de estructuras de gobernanza para desarrollar este tipo de transformación digital. Considerando este análisis, la sección siguiente propone una estrategia de apertura y uso intensivo de datos para el sector público.

4 ESTRATEGIA PARA LA IMPLEMENTACIÓN EFECTIVA DE UN GOBIERNO BASADO EN DATOS

En esta sección se propone una estrategia clara y factible para que las autoridades del Estado chileno construyan un modelo operativo de gobernanza de datos y política pública basada en datos. Se toma como referencia la relevancia, el marco conceptual, y la situación actual de nuestro Estado. Esta estrategia propone una respuesta a la pregunta: ¿cómo alcanzamos la situación deseada considerando las capacidades y contexto probable?

Con esta estrategia se puede derivar una estructura de producción del sistema de gobierno basado en datos. Aunque parezca evidente, la mayor parte de los intentos de cambio estratégico presentan fallas ya sea en el diseño estratégico o en su implementación. Por este motivo, el diseño estratégico debe ser realista respecto de la capacidad de implementación: no puede basarse en supuestos improbables, y por lo tanto debe considerar tanto las fortalezas como los desafíos existentes.

Entenderemos por estrategia la definición del cómo se logra implementar la visión, estableciendo las prioridades, determinando los recursos disponibles y necesarios, diagnosticando las oportunidades y obstáculos que operan en el contexto real, y dejando claro el foco en cada periodo de tiempo. Con ello, se minimizará la dispersión y los cambios en la estrategia por contingencias emergentes, o cambios de preferencias en el plan que generan altos costos para los equipos responsables. Así, la estrategia debe buscar converger todas las dimensiones de la transformación digital en el sector público (sección 2.4, Figura 5), donde el liderazgo político es crucial.

Para el caso, se quiere crear un sistema en el estado chileno que, por una parte:

- a) Desarrolle una capacidad de producir, gestionar y disponibilizar datos para análisis relacionados con políticas públicas y, que, por otra,

- b) Use significativamente esos datos para la optimización de la toma de decisiones, y el diseño, evaluación e implementación de políticas públicas.

Ambos objetivos se retroalimentan entre sí. Si existen buenos datos disponibles, se reducen los costos de transacción y por lo tanto habrá potencialmente más uso tanto dentro como fuera del Estado. Si ese uso es relevante, habrá más demanda por extender el modelo. Por lo tanto, se debe avanzar en ambos frentes. Si solo hay foco en desarrollar el sistema de producción, gestión y disponibilización, pero sin uso, el sistema caerá en el descrédito y tendrá menos apoyo político-financiero. Por otro lado, si solo usamos los datos disponibles, pero sin un sistema de producción y disponibilización, el desarrollo podría tener poco alcance por la frustración que implicará no tener datos para trabajar y un enorme costo de oportunidad.

En cuanto a la **capacidad de producir, gestionar y disponibilizar datos para análisis**, como se vio en la sección anterior, el impacto de los datos abiertos es limitado a la fecha, ámbito en el cual el indicador de la OCDE nos ubica en el antepenúltimo lugar solo superando a Letonia y Turquía (ver Figura 6). Esto hace que nuestro país se encuentre aún en niveles tempranos en el modelo de madurez de gobierno basado en datos (sección 3).

La causa es que, como se discutió en la sección 2.2, los datos abiertos no se producen con la intencionalidad de involucrar a los otros actores del ecosistema en responder preguntas estratégicas o disminuir los costos de transacción entre reparticiones públicas. Es decir, no siguen la cadena de generación de valor a partir de datos (sección 2.1) sino más bien responden principalmente a una lógica de rendición de cuentas.

Considerando estos objetivos y la situación actual, las siguientes secciones delinean una estrategia general que dan una pauta de actuación y, por otra, se desarrolla líneas de trabajo más específicas para su implementación.

4.1 DEFINICIONES CLAVE DE LA ESTRATEGIA

La definición principal de la estrategia es que, en un periodo de mediano plazo, el Estado chileno necesita ocupar intensivamente los datos para la toma de decisiones, diseñar e implementar mejor nuevas políticas públicas y/o optimizar las ya existentes. Un plazo que se estima razonable *a priori* es al **2022**, comenzando el 2018.

Una segunda definición es que la estrategia debe comenzar desarrollando las medidas más fáciles de implementar y que sean costo-efectivas al propósito. Se debe asegurar el logro evidente del uso intensivo de datos, de tal manera que las autoridades tengan retroalimentación positiva respecto de la política. Esto requiere desarrollar casos de éxito temprano (*quick-wins*) que permitan relevar prontamente la importancia de implementar un gobierno basado en datos.

Una tercera definición crítica respecto a la gobernanza y ejecución de este sistema es que a diferencia de algunas propuestas (Centro de Estudios Públicos, 2017), **no parece razonable concentrar todas las funciones en un solo lugar**. Primero, porque existen partes del sistema que necesariamente deben ser producidas de un modo distribuido, sin perjuicio de las coordinaciones tecnológicas y estandarizaciones necesarias. Y, en segundo lugar, porque apostar por una concentración del liderazgo en una sola estructura genera un alto riesgo de fracaso si esa unidad no logra implementar adecuadamente sus proyectos. Esto no quiere decir que no se deba fomentar un liderazgo claro y generar las especializaciones necesarias para empujar el sistema.

Finalmente, la cuarta definición es que debe fomentarse el uso de datos abiertos, transparentando todos aquellos datos no sensibles y previamente anonimizados según la normativa vigente. De esta forma, se disminuyen los costos de transacción dentro del mismo gobierno para el intercambio de información, y es fácil de acceder para la ciudadanía maximizando la oportunidad de que los datos sean utilizados. A excepción de los datos sensibles –que requieren de una definición operacional explícita– todos los datos debieran estar en modalidad abierta y además legibles por máquinas a través de políticas de apertura por diseño o por defecto (*open by design* u *open by default*). Por lo tanto, se requiere de instructivos y estándares sobre cuándo y cómo las distintas reparticiones publicarán sus datos.

4.2 COMPONENTES DEL SISTEMA DE GOBIERNO BASADO EN DATOS

La estrategia considera que, en Chile, el gobierno basado en datos tendrá dos grandes componentes: un sistema de producción, recolección y disponibilización de datos y otras unidades que usarán los datos para el análisis y toma de decisiones.

SISTEMA DE PRODUCCIÓN, RECOLECCIÓN Y DISPONIBILIZACIÓN DE DATOS

Respecto al **sistema de producción, recolección y disponibilización de datos**, que representa un modelo de gobernanza de datos para el sector público, se considera conveniente que se construya desde las capacidades tecnológicas y de estandarización que son responsabilidad de la recientemente creada División de Gobierno Digital dentro de SEGPRES. El sistema está pensado desde la lógica genérica de un servicio compartido estandarizado. Se espera que la División de Gobierno Digital avance hacia una estrategia que implemente y establezca gobernanza sobre la cadena de generación de valor de la información (sección 2.1) para la que la disponibilización de datos abiertos esté orientada a potenciar el ecosistema de generación de valor a partir de los datos (sección 2.3). Además de recolectar y disponibilizar datos de otras reparticiones públicas, SEGPRES puede producir una importante cantidad de datos para análisis a partir del programa [ChileAtiende](#).

USO DE ANALÍTICA PARA LA TOMA DE DECISIONES

No se considera conveniente que la División de Gobierno Digital asuma la responsabilidad única de liderar las iniciativas de análisis de datos. Esto porque para generar evidencia para la toma de decisiones y el diseño, implementación y evaluación de políticas públicas, lo primero que se debe tener presente **son las preguntas** que es necesario responder. Son cada una de las reparticiones públicas, dada la operación particular de su negocio, las que tienen la responsabilidad de producir esas preguntas, y por lo tanto debieran en el largo plazo producir también las respuestas.

Desarrollar este proceso de levantamiento de preguntas de otras reparticiones desde la División de Gobierno Digital de SEGPRES puede desfocalizarla y hacer que se aleje de sus propósitos y capacidades naturales. Esto no quiere decir que si se presentan oportunidades evidentes no se aprovechen (especialmente en el caso de ChileAtiende), pero en principio debieran ser otras unidades las encargadas de estas funciones. Estas últimas unidades pueden ser varias, y no tienen por qué responder a una lógica jerarquizada, ya que se desea que en el largo plazo las capacidades de analítica estén disponibles en todas las reparticiones que lo requieran.

Para que efectivamente el uso de analítica se implemente, también es necesario generar incentivos para que la formulación de políticas públicas tenga una sólida justificación basada en evidencia. Se

propone que, tal como actualmente ocurre en Nueva Zelanda, el financiamiento de programas sociales siga dos caminos:

- Si la política pública usa evidencia justificada mediante un análisis robusto de datos del país, entonces no se establece un límite a priori del monto que se puede destinar al programa y se integra esta solicitud al proceso presupuestario normal, donde se determinará el monto del presupuesto que obtendrá de acuerdo con la disponibilidad de recursos y otras consideraciones.
- En caso en que no existan los análisis, cada programa puede a lo más recibir un monto máximo fijado previamente.

Esto permite avanzar en **desarrollar estándares de calidad** en el mecanismo de evaluación *Ex ante* para programas nuevos y la ampliación de los existentes. Esto debe ser válido también para el gasto que se incorpore a partir de la aprobación de nuevas leyes, donde —proponemos que— la evidencia deba suministrarse como parte de alguna de las siguientes instancias: 1) los informes financieros que genera DIPRES; o 2) los informes de productividad que elaborarán los distintos Ministerios según el proyecto de ley actualmente en tramitación el Congreso Nacional.

El propósito de usar estas dos instancias es responder preguntas básicas respecto de la existencia de los insumos necesarios para generar evidencia, guiadas por la cadena de generación de valor: ¿existe un modelo de datos? ¿cuáles son los mecanismos de captura y gestión de esos datos? ¿con qué propósito se utilizarán? ¿qué estándares se utilizarán? ¿cuáles datos se transformarán en abiertos?, entre otras.

PRINCIPALES ACTORES QUE DEBEN IMPULSAR EL GOBIERNO BASADO EN DATOS

En principio, se identifican dos grandes actores en el sistema, estos son: la División de Gobierno Digital y el Ministerio de Hacienda en conjunto con la Dirección de Presupuesto (DIPRES).

La **División de Gobierno Digital**, como ya se señaló, cuenta con definiciones, mandato, prácticas, y capacidades en desarrollo para la construcción del sistema de producción, recolección y disponibilización de datos. Dentro de la dicha división, debiera localizarse la función del *Chief Data Officer* (CDO), figura encargada de implementar la gobernanza de la cadena de generación de valor a partir de los datos (sección 2.1) en sus tres dimensiones: disponibilidad, accesibilidad (haciendo énfasis en la **calidad**) y re-utilización de datos. En la dimensión de re-utilización de datos, es clave que el CDO, en función de las necesidades de analítica del Estado, coordine la intencionalidad y priorización de publicación de datos abiertos que alimenten un **ecosistema de generación de valor** que involucre y articule a los actores clave en el uso intensivo de datos: los propios funcionarios públicos, la ciudadanía (que incluye la academia y las organizaciones de la sociedad civil) y el sector privado (sección 2.2). Se deben entonces definir actividades específicas que permitan consolidar este ecosistema y desarrollar una cultura de apertura y uso inteligente de datos tanto dentro de las instituciones públicas como fuera de ellas.

El **Ministerio de Hacienda y la Dirección de Presupuesto** (DIPRES). En el caso del Ministerio de Hacienda, debiera liderar y apoyar el desarrollo de los proyectos en **otras reparticiones** que sean necesarios para desarrollar los casos de éxito temprano, sin asumir funciones operativas permanentes. Asimismo, dentro de DIPRES debiera localizarse la primera unidad de analistas de datos de nivel estratégico, ya que esto complementa directamente la labor de dicha dirección.

En este último caso las razones para privilegiar a DIPRES como unidad de análisis de datos son las siguientes:

1. DIPRES debe tomar decisiones anuales presupuestarias del orden de **65.000 millones de dólares** y diseña sistemas de largo plazo. Por lo tanto, es fundamental robustecer sus capacidades de análisis que tendrán un importante retorno a la inversión de corto plazo.
2. DIPRES, por su posición en términos presupuestarios y normativos, puede introducir cambios relativamente rápido en varios procesos (como por ejemplo en inversiones tecnológicas, diseños de programas, entre otras) ya sea para fomentar la analítica o para implementar recomendaciones en planes y programas producto de sus propios análisis.
3. Dado que DIPRES es una estructura transversal del Estado, tiene visibilidad simultánea de distintas temáticas (por ejemplo, Salud, Educación, Desarrollo Social y Seguridad pública) y por lo tanto es un punto neurálgico donde pueden integrarse datos desde distintas fuentes que permita generar evidencia para mejorar el trabajo intersectorial.
4. Finalmente, DIPRES cuenta con una fuerte tradición en análisis datos con metodologías tradicionales asentadas en la unidad de estudio y de gestión públicas, y validadas ante el resto del Estado.

Cabe destacar que esta unidad en DIPRES no tiene que ser la única unidad que realice análisis del Estado. De hecho, el rol de Hacienda desde el Programa de Modernización del Sector Público ha sido precisamente fomentar el desarrollo de analítica en otras Unidades para generar casos de éxito temprano. Sin embargo, mejorar las capacidades analíticas de DIPRES es un elemento esencial para impulsar el gobierno basado en datos, por las razones anteriormente explicadas.

IMPLEMENTACIÓN Y ACOMPAÑAMIENTO DE LA ESTRATEGIA

La implementación de la estrategia debe ser acompañada por organismos internacionales competentes que aseguren conocimiento continuo de las mejores prácticas, como, por ejemplo, la **Organización para la Cooperación y el Desarrollo Económico (OCDE)**. La OCDE ya ha estado colaborando previamente en los diagnósticos y estrategias de gobierno digital (OECD, 2016) y capacidades de los funcionarios públicos (OECD, 2017c), y de su experiencia provienen muchos de los modelos anteriormente discutidos en el marco conceptual como la cadena de generación de valor a partir de los datos y el ecosistema de generación de valor a partir de los datos (OECD, 2017a). Además, realiza diagnósticos sobre el uso y apertura de datos en el sector público en diferentes países pertenecientes a la organización (OECD, 2017b).

Otro organismo multilateral que está potenciando el uso de analítica dentro de los gobiernos de América Latina es el **Banco Interamericano de Desarrollo (BID)**. El BID a través de su sector de instituciones para el desarrollo, promueve una serie de proyectos y estudios para sistematizar experiencias y generar recomendaciones para los distintos países (Farias et al., 2016; Goldsmith et al., 2016; Rodríguez et al., 2017; Tomar et al., 2016; Wiseman, 2015, 2017a). Además, en este momento financia el Programa de Modernización del Sector Público del Ministerio de Hacienda que su vez financió el desarrollo de este estudio.

Otros gobiernos con experiencia en el tema son **Nueva Zelanda** (Mansell, Laking, Matheson, & Light, 2017), el **Reino Unido** (Government of United Kingdom, 2017; Tomar et al., 2016), y algunas **ciudades y estados de EE.UU** (Athey, 2017; Goldsmith et al., 2016; Tomar et al., 2016; Wiseman, 2015, 2016, 2017a).

Es necesario que durante la implementación de esta estrategia se potencie el ecosistema de generación de valor fomentando el trabajo con entidades nacionales e internacionales tanto del sector privado como la ciudadanía. Especialmente importante es el rol de la academia y organizaciones no gubernamentales que han impulsado el uso de analítica dentro y fuera de Chile. Los centros de investigación de las Universidades más importantes del país han contribuido desde hace años a este rol al que se han sumado organizaciones como por ejemplo, [Fundación Ciudadano Inteligente](#) y el [Observatorio del Gasto Fiscal](#).

En el ámbito académico internacional, podemos nombrar iniciativa que explícitamente se encargan de apoyar el uso de analítica en el sector público como por ejemplo, [The Center for Government Excellence](#) de la Universidad Johns Hopkins, [Data Science for Good](#), de la Universidad de Chicago, el [Ash Center](#) de la Universidad de Harvard, y el [GovLab](#) de la Universidad de Nueva York. Si bien los insumos para los análisis de datos debieran corresponder principalmente a datos del sector público, hay que prestar atención a las oportunidades que ofrecen otras fuentes abiertas como por ejemplo de iniciativas como *OpenStreetMap*.

Por último, la implementación de la estrategia debiera identificar y promover la coordinación mediante mecanismos efectivos, liderados por el binomio División de Gobierno Digital – Ministerio de Hacienda con todos aquellos actores que por razones técnicas o jurídicas tienen o podrían cumplir un rol clave. Por ejemplo, en la ley actualmente en discusión en el congreso se le entregan al INE facultades importantes para solicitar datos administrativos para fines estadísticos. Por su parte, el Registro Civil tiene actualmente obligación legal de mantener los catálogos de datos del Estado, además de administrar la identidad de los ciudadanos, elemento clave para la coordinación de información y que constituye una de las fuentes de datos más grande del Estado. Asimismo, para cumplir sus funciones el Ministerio de Desarrollo Social (MDS) tiene la facultad de pedir datos a otros organismos del Estado, lo que le otorga un potencial de almacenamiento y procesamiento de datos de distintas fuentes como ningún otro Ministerio del Gobierno. Esta coordinación debiera expresarse en un consejo consultivo/asesor del sistema.

4.3 ESPECIFICACIONES DE LA ESTRATEGIA

ESQUEMA GENERAL

En la Figura 9 se muestra en los principales componentes del sistema de gobierno basado en datos para Chile. En esta visión de largo plazo, la cúspide tiene la función de gobernanza. En el lado izquierdo del cuadro se encuentra el sistema tecnológico-organizacional-normativo de producción, gestión y disponibilización de datos, cuyo liderazgo le corresponde a la División de Gobierno Digital. A la derecha de la figura se encuentra el sistema de análisis de datos, que es un sistema que coordina a los distintos actores del ecosistema de generación de valor.

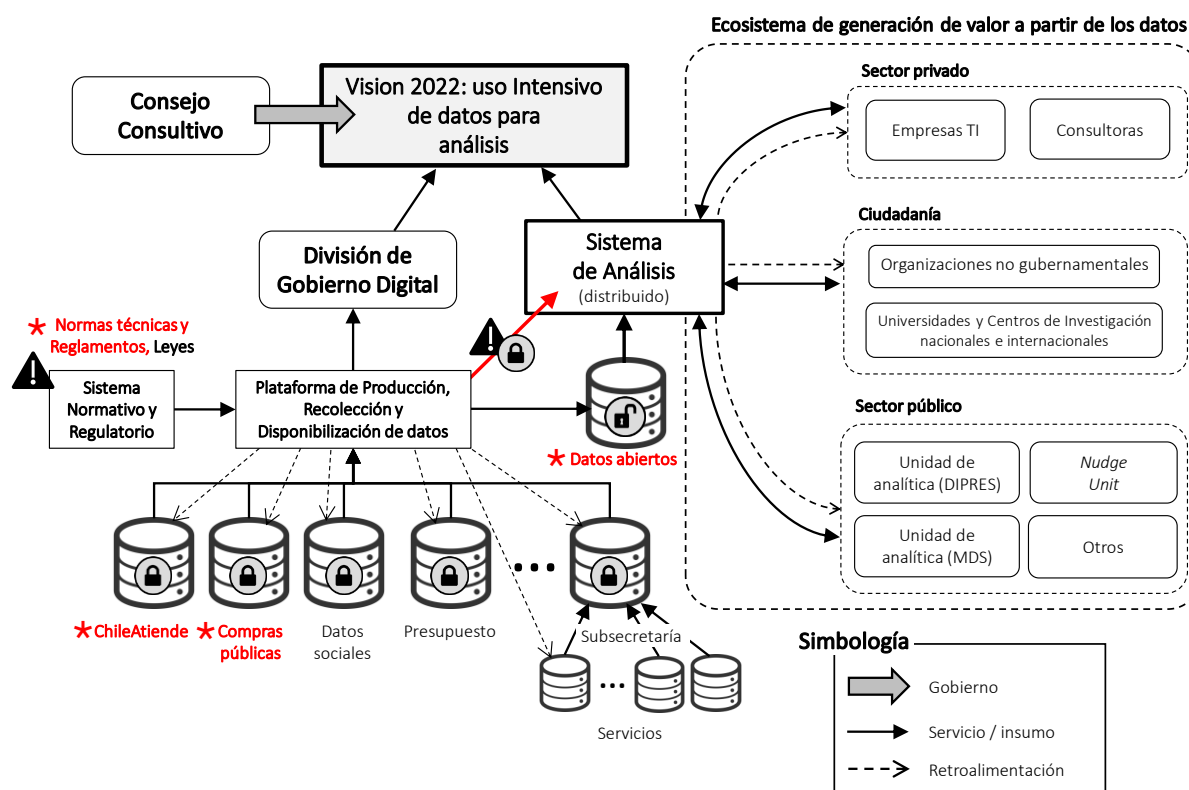


Figura 9: Componentes del sistema de gobierno basado en datos para Chile.

Como aspecto central, la gobernanza de datos pública debiera ser liderada por la **División de Gobierno Digital**, con un CDO encargado de implementar la gobernanza de la cadena de generación de valor a partir de datos, coordinando la intencionalidad y priorización de publicación de datos abiertos que alimenten el ecosistema de la generación de valor y la participación de la ciudadanía y el sector privado.

No obstante, el análisis de datos no debiera centralizarse, ni ser responsabilidad exclusiva de la División de Gobierno Digital ya que como se mencionó anteriormente, se requiere de cierta especialización para poder formular las preguntas que requieren respuestas con procesos de analítica, y esa experticia está necesariamente distribuida en las distintas unidades del Estado. Para ello se propone desarrollar creando una masa crítica de analítica dentro del Estado, partiendo con una unidad analítica en DIPRES como se discutió en la sección 4.2.

Sin embargo, se hace necesaria la coordinación de ciertos actores del sector público que son grandes productores y/o consumidores de datos o tienen ciertas facultades administrativas y legales entorno a éstos (sección 4.2). Por lo tanto, para coordinar a estos actores se recomienda crear un **consejo consultivo** compuesto, en principio, por:

- La **División de Gobierno Digital**, encargada responsable de la gobernanza del sistema.
- El **Ministerio de Hacienda**, por ser uno de los principales usuarios de los datos a través de DIPRES y como por la capacidad de asegurar los recursos estratégicos para la implementación y sustentabilidad de la estrategia.

- El **Instituto Nacional de Estadísticas (INE)**, por sus facultades normativas para recolectar datos administrativos, por ser una entidad con un marco institucional que asegura la confianza en el manejo de bases de datos, y por ser un usuario privilegiado de los datos.
- El **Ministerio del Interior**, por sus actuales funciones en seguridad informática (que eventualmente podrían ser en el futuro competencias de la División de Gobierno Digital).
- El **Ministerio de Desarrollo Social (MDS)** por el desarrollo del sistema de información social como por necesidad y capacidad de procesamiento de dicha información.
- Eventualmente, el **Registro Civil** debiera ser considerado por sus funciones de mantención del catálogo de bases de datos del Estado y por contar con las bases más extensas, sus atribuciones y capacidades en materia de identidad digital, entre otras.

Este consejo debiera tener funciones al menos de coordinación y asesoría, y además permitiría ser un catalizador de la experiencia que se recoja de otras unidades hacia el interior de esas reparticiones públicas.

SISTEMA DE PRODUCCIÓN, RECOLECCIÓN Y DISPONIBILIZACIÓN DE DATOS

En cuanto al **sistema de producción, recolección y disponibilización de datos**, este lo conceptualizamos como un servicio compartido para todas reparticiones públicas para el manejo de datos del Estado, abiertos o no. Este sistema debe contener en sí las estandarizaciones necesarias, debe facilitar la recolección y disponibilización de datos, minimizando los costos de transacción. Este sistema deberá proveer diferentes servicios, tales como: estandarización y procesamiento para *open data*, consultas anónimas, procesamiento *online*, cruces de datasets, y otros mecanismos que se ajusten a la demanda respetando las restricciones normativas.

Asimismo, el sistema debiera facilitar la disponibilización de datos por parte de los servicios productores. Lo ideal es que la captura del dato sea desde la fuente original de las operaciones de los servicios y el hecho que sea la División de Gobierno Digital la entidad además encargada de la interoperación de datos, lo facilita. Sin embargo, debieran seguirse estrategias diversificadas según las diferentes realidades de desarrollo tecnológico de los servicios públicos. Por ejemplo, el caso servicios como ChileCompra, el Servicio de Impuestos Internos (SII) y ChileAtiende que cuentan con sistemas tecnológicos avanzados de *backoffice* es diferente al de otros servicios que no tienen sistemas en operación y que por tanto requieren de proyectos para el desarrollo de dicho *backoffice*. En este último caso, una línea de acción de mediano y largo plazo es que sean las Subsecretarías de los distintos ministerios, las que lideren y fomenten el desarrollo de estos sistemas para sus sectores. En cualquier caso, se debe evitar el enfoque de un mega sistema homogéneo para todas las unidades. El Consejo consultivo podrá contribuir a fijar las prioridades y el Programa de Modernización podrá ayudar a financiar los proyectos que se requieran según estos lineamientos.

Otro aspecto particular que relevar en la puesta en marcha del sistema de producción, recolección y disponibilización de datos, es el uso de los **datos abiertos** como herramienta para reducir los costos de transacción entre diferentes entidades públicas entre las cuales muchas veces no es fácil compartir datos, ya sea por exceso de celo legal o falta de conocimiento de las normas que facultan la interoperabilidad de datos del Estado, o por la baja calidad de los datos que gestionan. El primer aspecto a resolver es **definir una política de datos abiertos** para el Estado. Esto significa definir roles, procedimientos y recursos tanto a nivel sectorial como central en la División de Gobierno Digital que faciliten el flujo de datos hacia el sistema centralizado de recolección y disponibilización, aspectos

que no han sido abordados totalmente a la fecha (Gonzalez-Zapata, 2016; González-Zapata & Heeks, 2017). No obstante, lo anterior, es necesario también zanjar un marco normativo que facilite el uso genérico de datos como insumo para la analítica y no exclusivamente para los fines iniciales de la recolección, resguardando debidamente los derechos de los ciudadanos. Asimismo, en caso de ser necesario realizar análisis sobre datos no abiertos y que se requiera la participación de otras entidades fuera del Estado (por ejemplo, por sus competencias y experiencia en este ámbito) se deben tomar los resguardos de protección de los datos, **previo** al traspaso de estos especialmente referente a qué se pueden hacer con los datos cuando haya terminado el análisis de los mismos y la propiedad de éstos, que no debe cederse ni permitirse que aquellos que desarrollaron los análisis los usen para otros fines.

Sin embargo, para que la estrategia sea exitosa es necesario abordar el problema de la **calidad** de los datos que se disponibilizan. La calidad de los datos es muy diversa entre las reparticiones públicas. Hay algunas que no tienen datos de calidad, otras que registran algo (pero que en la práctica no sirven o incluso pueden inducir a error), otras que registran (pero con fallas), hasta el otro extremo del espectro, donde hay un registro completo y acusioso.

La publicación de datos en formato abierto también es un mecanismo para que se audite externamente la calidad de los mismos, principalmente por la sociedad civil, como propone el ecosistema de generación de valor a partir de los datos (sección 2.2). Con dicha publicación se espera que en el mediano o largo mejore la calidad, pero en el intertanto pueden datos con errores importantes, con el consecuente impacto: desconfianza, si es usado por sociedad civil o malas decisiones, si es directamente utilizado por el sector público. Pero seguir un esquema de "publicar sólo cuando esté bueno" simplemente romperá la posibilidad de que el ecosistema de generación de valor apoye la mejora de los datos.

Como estrategia solución, se sugiere publicar los datos “tal como están” (del inglés ‘as is’), tanto en formato abierto como solo disponibles para el Estado, junto con advertencias de distinto nivel. Un primer nivel, corresponde un análisis básico —que puede provenir de la División de Gobierno Digital— para alertar sobre potenciales datos mal registrados. En segundo nivel, es una advertencia de la propia institución que disponibilizó los datos con respecto a algunos casos¹⁴, cuyas anotaciones deben ser parte integral del *dataset*, como parte de la metadata asociada. Asimismo, las correcciones a los *dataset* (provenientes de advertencias de nivel 1 y 2) deben quedar registradas. Por lo tanto, cuando se disponibilicen los *datasets*, además de los datos vendrá documentación respecto a las metodologías de recolección, alcances, y advertencias respectivas.

Sin establecer una estrategia que permita que los datos se expongan dejando claro las consecuencias cuando existan errores, es muy difícil que se disponibilicen, lo que es fundamental para la generación de valor y la toma de decisiones. Un mecanismo como el propuesto permite que los datos se expongan, y a través de la trazabilidad de los cambios, las advertencias y la retroalimentación del ecosistema de generación de valor se pueden, por ejemplo, comprometer fechas de reparación. Por lo tanto, en el corto plazo, una de las labores del CDO debiera concentrarse en un plan que permita generar datos de calidad e implementar la gobernanza de la generación de valor en aquellas

¹⁴ Por ejemplo, si en el mes de marzo una institución envía toda un *dataset*, varias veces, al registro civil para validación cruzada de ciertos datos, esto queda registrado como un “tramite” en canal digital, disparando el respectivo indicador, sin que sea un aumento orgánico de los trámites. Este tipo de casos debieran publicarse como advertencia en el *dataset* de transacciones de la institución que se publique.

reparticiones que requieran responder preguntas estratégicas y urgentes, mientras que el resto de las reparticiones opera en la lógica de publicación “tal como están”, con las medidas del caso.

Adicionalmente, y en el muy corto plazo, existe una oportunidad de mejorar el portal de datos abiertos (datos.gob.cl) a través de la integración con la plataforma de integración y visualización de datos abiertos DataChile (www.datachile.io). En este portal los datos abiertos se integran para crear visualizaciones interactivas que facilitan la navegación y análisis de los datos públicos que contiene, en vez de simplemente ser un repositorio de archivos tradicional como los construidos en plataformas de código abierto como CKAN (ckan.org). El portal DataChile incorporará también una herramienta para combinar distintas fuentes de datos, lo que facilitará además del proceso de búsqueda y combinación de datos para la analítica. Sin embargo, DataChile actualmente depende de la Dirección General de Relaciones Económicas Internacionales (DIRECON), del Ministerio de Relaciones Exteriores, mandante para el Programa de Bienes Públicos de CORFO, y por lo tanto requiere ser trasladado a la División de Gobierno Digital dada la naturaleza de DataChile como servicio compartido de integración y visualización de datos, y además para que exista una mejor integración con la disponibilidad de datos abiertos que promueve dicha División.

Se sugiere, además, considerar experiencias exitosas de apertura de datos públicos en el extranjero para analizar estrategias de datos abiertos que sean consistentes con esta propuesta. En particular, resulta de interés el caso de la plataforma de datos abiertos de Francia (www.data.gouv.fr), que se ha transformado en un sistema de interacción entre ciudadanos y el Estado francés. A diferencia de la mayoría de los portales de datos abiertos centralizados – que consisten en una relación unidireccional de descarga de *datasets* – el portal de datos abiertos de Francia dispone dos interesantes mecanismos para que los ciudadanos interactúen con el Estado. En primer lugar, los ciudadanos pueden subir sus propios *datasets*, a partir de la recolección de datos que ellos hayan realizado o el cruce de *datasets* existentes para obtener datos más refinados. En segundo lugar, los ciudadanos pueden publicar, en formato blog, sus proyectos de reutilización de datos, lo que permite dar visibilidad al trabajo que realiza la comunidad de usuarios con datos abiertos (tanto dentro como fuera del sector público) y generar conciencia dentro de las instituciones públicas del valor público asociado a los datos que producen. Ambos mecanismos son ejemplos interesantes de cómo fomentar y fortalecer el ecosistema de apertura y reutilización de datos públicos, lo que tiene directo impacto en el valor público generado a partir una estrategia de *data-driven public sector* como la sugerida en este estudio.

USO DE ANALÍTICA PARA LA TOMA DE DECISIONES

Por el lado del **sistema de análisis de datos**, debieran existir diversas unidades que cumplan esa función. Para el corto plazo, se recomienda partir con una unidad de alto nivel en DIPRES. De todas maneras, se debe promover el trabajo en red y prácticas que permitan compartir el conocimiento. En la Figura 9 se destacan los **puntos mínimos y prioritarios (en rojo)** que se considera debieran ser el foco de la acción en el corto plazo.

Algunos de esos focos debieran ser por un lado **ChileAtiende**, lo que permite entender como el Estado se relaciona con el ciudadano a través de los diferentes servicios que este les provee y cómo se podrían hacer más costo-efectivos, y por otro las **Compras Públicas**, que genera una información tremendamente valiosa para optimizar el gasto del Estado, desde la detección de colusiones en la presentación de ofertas y análisis históricos de evolución de precios, hasta predicción de adquisición de productos y servicios por parte de las distintas reparticiones públicas que permitan generar volúmenes y condiciones de compra más convenientes, entre otras.

En relación con los **perfiles profesionales** necesarios, se puede delinear que en el caso de la División de Gobierno Digital debiera componerse por ingenieros civiles expertos en tecnologías, a nivel estratégico, de arquitecturas tecnológicas, más ingenieros y programadores especialistas. Asimismo, esta unidad debiera contar con especialistas jurídicos en temas de protección y derecho administrativo. En el caso de las unidades especializadas de datos, además de los especialistas temáticos según sea el caso de las políticas públicas objeto de los análisis, debieran incorporarse científicos de datos, ingenieros civiles y estadísticos con manejo avanzado en metodologías nuevas en análisis de datos masivos (*'Big Data'*), *Machine Learning*, entre otras técnicas.

En materia de **socios estratégicos**, en una primera instancia se observa que la clave está en los mecanismos de coordinación de la División de Gobierno Digital, el Programa de Modernización del Sector Público del Ministerio de Hacienda y DIPRES. Difícilmente una de estas unidades tendrá todas las capacidades para enfrentar este desafío. Las debilidades de alguna unidad podrán ser compensadas por las otras partes. Otro actor que podría ser importante incorporar es el **Laboratorio de Gobierno**, que puede tener un rol clave en la estrategia de adopción por parte de los funcionarios del sector público. De cualquier forma, deben tener funciones claramente diferenciadas, pero siempre actuando de modo coordinado.

En relación con las **alianzas estratégicas**, se debería partir el primer año con el desarrollo de acuerdos de cooperación y estudios *ad hoc* con la OCDE en esta materia y con otros organismos especializados de altos nivel, ya sean públicos o centros de investigación de alto nivel. En esta línea, debiera desarrollarse acuerdos de cooperación con países avanzados en la materia como Nueva Zelanda, el Reino Unido y Francia. Las relaciones debieran diseñarse para un mediano plazo a lo menos.

4.4 INICIATIVAS PRIORITARIAS EN EL CORTO Y MEDIANO PLAZO

No se puede hacer todo al mismo tiempo, y es necesario asegurar resultados relativamente rápidos para permitir a las autoridades visibilizar de la viabilidad y valor de esta política. En la Figura 10 se muestra la secuencia de iniciativas desde el 2018 al 2022. En la parte superior se indican los principales hitos de las iniciativas que debe abordar la División de Gobierno Digital según la estrategia definida anteriormente. Del mismo modo, en la parte inferior, se indican las tareas del Ministerio de Hacienda y la DIPRES. También se incluyen en esta hoja de ruta, estudios que se estima necesario realizar para levantar insumos e información importantes para la implementación y elaboración de planes más detallados. Para mayores detalles de estos estudios referirse al “Documento con líneas de estudio/trabajo para el Programa de Modernización del Sector Público”.

Para el **2018** se debiera partir con dos acciones clave: el inicio de la función de *Chief Data Officer* (CDO) en la División de Gobierno Digital y la instalación de la unidad de analítica en DIPRES:

En el caso del CDO, dicho profesional a cargo debiera ser elegido en el marco de un concurso de alta dirección pública y debiera apostar por reclutar a la mejor capacidad técnica-estratégica disponible para el propósito. El objetivo es la instalación de la unidad, diseño de un plan estratégico e identificación de proyectos para desarrollar. El mismo año, el equipo del CDO debiera constituir el consejo consultivo e iniciar los acuerdos de cooperación.

En el caso de DIPRES debieran incorporarse al equipo 2 o 3 expertos que den el puntapié inicial a una unidad especializada en técnicas avanzadas de datos. En paralelo debiera enfatizarse la apertura de datos abiertos en sectores relevantes, considerando los ejemplos internacionales exitosos mencionados previamente que permitan fortalecer el ecosistema de reutilización de datos.

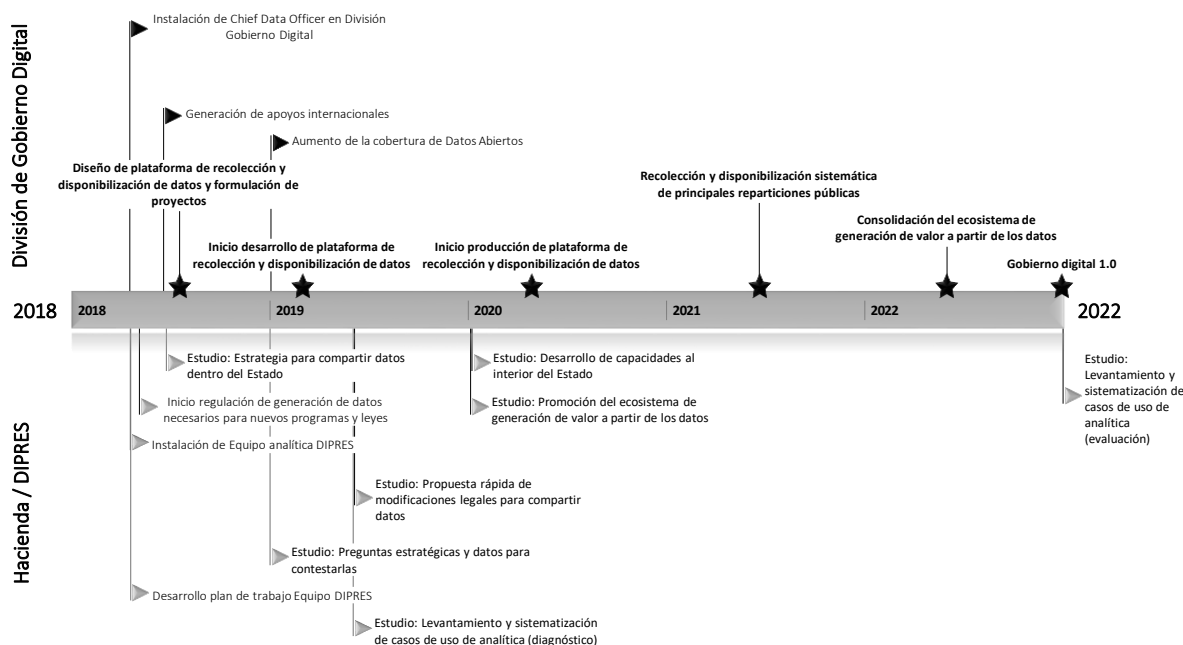


Figura 10: Hoja de ruta 2018 — 2022 de la estrategia de desarrollo de un gobierno basado en datos.

El **2019** debiera partirse con la implementación del sistema de producción y disponibilización de datos y las primeras regulaciones y estandarizaciones del sistema. Asimismo, la unidad de análisis de datos en DIPRES debiera mostrar los principales resultados relevantes. El **2020**, el sistema de producción y disponibilización de datos debiera entrar en fase de producción, para que las distintas reparticiones públicas empiecen a disponibilizar sus datos empujadas por las preguntas estratégicas que se deben responder, y esa información —en una cantidad significativa, esté disponible el **2021** y se pueda desarrollar el ecosistema de generación de valor para **2022**. Esto permitirá lograr como objetivo tener una primera versión de un gobierno que use intensivamente los datos para la toma de decisiones para finales de dicho año.

4.5 ESTRATEGIA A TRAVÉS DEL MODELO DE TRANSFORMACIÓN DIGITAL

Dado lo discutido previamente en la sección 3.5 , es fundamental que la estrategia considere las tres dimensiones necesarias para conseguir la transformación digital en la apertura y uso intensivo de datos, estas son: liderazgo político, la gobernanza y capacidad institucional y los recursos.

Respecto al **liderazgo político**, es fundamental que dado esta estrategia —que incluye un *roadmap* y responsabilidades claras en distintos actores— la autoridad se comprometa a llevarlo a cabo asignándole una alta prioridad y haciéndole un seguimiento y evaluación sistemáticas para asegurar que se cumplan los hitos críticos de la estrategia. Es importante que la autoridad también empodere al CDO respecto de sus funciones y responsabilidades como actor de transformación transversal, al mismo tiempo que releve la importancia del uso intensivo de datos como parte esencial de la gestión pública del Estado. Este enfoque permitirá contar un suficiente respaldo político para movilizar una agenda tanto de gobernanza como de uso de datos públicos.

Respecto de la **gobernanza y capacidad institucional**, la estrategia establece un modelo de gobernanza a través de un consejo consultivo y la especialización de la División de Gobierno Digital en el ámbito de la recolección y gestión de datos de acuerdo con la cadena de valor de la información. Este modelo establece un objetivo principal: producir datos para responder las preguntas estratégicas tanto de los diferentes servicios y como de la gestión central del Estado. También se destaca el rol de gestión centralizada sobre los datos otorgado a esta División, dado que el sistema de análisis queda distribuido y se apoya en desarrollar un ecosistema de generación de valor a partir de los datos a partir de este modelo de gobernanza de datos. Además, la estrategia define la generación de capacidades institucionales en otras reparticiones que puedan generar beneficios inmediatos del uso de analítica de datos, como ejemplifica la propuesta de unidad de análisis en DIPRES. La división de roles respecto de la implementación de esta estrategia – la gobernanza de datos en la División de Gobierno Digital y la unidad de análisis de datos inicialmente en DIPRES – permite delinear responsabilidades y actividades específicas para cada institución, lo que se considera crítico para el éxito de esta estrategia debido a que la labor de la División de disponibilizar los datos para producir analítica requiere de un significativo esfuerzo de coordinación entre los diferentes servicios públicos.

Consistente con esta priorización, deben existir los **recursos** necesarios para que la estrategia se implemente según el plan propuesto, es decir para que esta División pueda de forma efectiva desarrollar un modelo centralizado de gobernanza de datos, a la vez que se puedan generar las capacidades institucionales en DIPRES y otras instituciones públicas para el uso intensivo de datos. Es importante además entender esta estrategia con una perspectiva de largo plazo y que no se evalúe en función de resultados inmediatos.

Agradecimientos

Quisiéramos agradecer el feedback provisto a este estudio por **Barbara-Chiara Ubaldi**, Directora de Gobierno Digital y Datos Abiertos, Secretaría de Gobernanza Pública, OCDE; y **Jane Wiseman**, Directora del Institute for Excellence Government y fellow del Innovations in Government Fellow, Ash Center for Democratic Governance and Innovation, Harvard Kennedy School. Sus comentarios expertos han sido una valiosa contribución para el desarrollo teórico, analítico y estratégico de este estudio. Sin embargo, todos los resultados, análisis y propuestas realizadas en este estudio son de exclusiva responsabilidad de los autores, y no necesariamente representan su visión ni la de sus instituciones.

REFERENCIAS

- Anselin, L. (1995). Local Indicators of Spatial Association—LISA. *Geographical Analysis*, 27(2), 93–115.
- Athey, S. (2017). Beyond prediction: Using big data for policy problems. *Science*, 355(6324), 483–485.
- Barredo, J. I., Kasanko, M., McCormick, N., & Lavallo, C. (2003). Modelling dynamic spatial processes: simulation of urban future scenarios through cellular automata. *Landscape and Urban Planning*, 64(3), 145–160.
- Bienkowski, M., Feng, M., & Means, B. (2012). *Enhancing Teaching and Learning through Educational Data Mining and Learning Analytics*. Washington: U.S. Department of Education, Office of Educational Technology.
- Castellón González, P., & Velásquez, J. D. (2013). Characterization and detection of taxpayers with false invoices using data mining techniques. *Expert Systems with Applications*, 40(5), 1427–1436.
- Centro de Estudios Públicos. (2017). *Un Estado para la Ciudadanía, Informe de la Comisión de Modernización del Estado*. Santiago: Centro de Estudios Públicos.
- Centro de Estudios sobre Desarrollo Económico. (2014). *Informe mensual sobre el soporte técnico y avance del contrato para garantizar la alimentación, consolidación, validación y uso de la información del SPADIES*. Bogotá: Universidad de los Andes.
- Corduneanu-Huci, C., Hamilton, A., & Ferrer, I. M. (2012). *Understanding policy change : how to apply political economy concepts in practice* (No. 73825) (pp. 1–379). New York: The World Bank.
- Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to Algorithms, Third Edition* (3rd ed.). Boston: The MIT Press.
- Cortana Intelligence and ML Blog Team. (2015). ML Predicts School Dropout Risk & Boosts Graduation Rates - Machine Learning - Site Home - TechNet Blogs. Recuperado 30 de diciembre de 2017, a partir de <http://blogs.technet.com/b/machinelearning/archive/2015/06/04/ml-predicts-high-school-dropout-risk-boosts-graduation-rates.aspx>
- Cortez, P., & Silva, A. (2008). Using data mining to predict secondary school student performance. En *Proceedings of 5th Annual Future Business Technology Conference*. (Vol. EUROSIS, pp. 5–12). Porto. Recuperado a partir de <http://hdl.handle.net/1822/8024>
- Curtatone, J. A. (2014). Open data only the beginning of the discussion. Recuperado 21 de diciembre de 2017, a partir de <http://www.thesomervilletimes.com/archives/47394>
- Dekker, G., Pechenizkiy, M., & Vleeshouwers, J. (2009). Predicting Students Drop Out: A Case Study. En *Computers, Environment and Urban Systems* (pp. 41–50).
- Development Gateway. (2017). *Informe final para plan de mejora de datos abiertos de ChileCompra*. Washington DC: Development Gateway.
- Escobar, C., & Lolas, F. (2015). *Desarrollo de un sistema prototipo para la detección temprana de la deserción escolar en escuelas públicas chilenas* (Memoria de Título). Universidad Adolfo Ibáñez, Santiago de Chile.
- Farias, P., Goldsmith, S., Flumian, M., Mendoza, G., Wiseman, J., Porrúa, M., ... Zanabria, G. (2016). *Gobiernos que sirven: Innovaciones que están mejorando la entrega de servicios a los ciudadanos*. Washington: Banco Interamericano de Desarrollo.
- Friedman, V. (2008). Data Visualization and Infographics | Smashing Magazine. Recuperado 18 de junio de 2014, a partir de <http://www.smashingmagazine.com/2008/01/14/monday-inspiration-data-visualization-and-infographics/>

- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big Data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
- Gobierno de Chile. (2012). Instructivo Presidencial N° 005 - 2012 sobre Gobierno Abierto. Santiago: Gobierno de Chile.
- Goldsmith, S., Crawford, S., & Weinryb Grohsgal, B. (2016). *Innovations in public service delivery - Issue No. 4: predictive analytics: driving improvements using data* (Documento para discusión No. IDB-DP-440). Washington: Inter-American Development Bank.
- Gonzalez-Zapata, F., & Heeks, R. (2015). The multiple meanings of open government data: Understanding different stakeholders and their perspectives. *Government Information Quarterly*, 32(4), 441-452.
- Gonzalez-Zapata, F. (2016). *Regional Report for Latin America - Open Data Barometer 3rd Edition*. Washington D.C.: The World Wide Web Foundation.
- Gonzalez-Zapata, F., & Heeks, R. (2017). The challenges of institutionalising open government data (OGD): A historical perspective of Chile's OGD initiative and digital government institutions. En *The Social Dynamics of Open Data*. Cape Town, South Africa: African Minds Publishing.
- Government of United Kingdom. (2017). Government Transformation Strategy: better use of data - GOV.UK. Recuperado 8 de enero de 2018, a partir de <https://www.gov.uk/government/publications/government-transformation-strategy-2017-to-2020/government-transformation-strategy-better-use-of-data>
- Hamel, M. B., Toth, M., Legedza, A., & Rosen, M. P. (2008). Joint replacement surgery in elderly patients with severe osteoarthritis of the hip or knee: decision-making, postoperative recovery, and clinical outcomes. *Archives of Internal Medicine*, 168(13), 1430–1440.
- Harvard Kennedy School. (2017). Operational Excellence in Government. Recuperado a partir de <http://www.innovations.harvard.edu/opex>.
- Hernández, M., Hong, L., Frias-Martinez, V., Whitby, A., & Frias-Martinez, E. (2017). *Estimating Poverty Using Cell Phone Data: Evidence from Guatemala*. New York: The World Bank.
- Hilbert, M., & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332(6025), 60–65.
- IBM. (2015). What is a Data Scientist? – Bringing big data to the enterprise. Recuperado 31 de diciembre de 2017, a partir de http://www-01.ibm.com/software/data/infosphere/data-scientist/?cm_mc_uid=23497733502814413056500&cm_mc_sid_50200000=1441305650
- Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301), 790–794.
- Jolliffe, I. T. (2002). *Principal Component Analysis* (2nd ed.). New York: Springer-Verlag.
- Joseph, A. E., & Bantock, P. R. (1982). Measuring potential physical accessibility to general practitioners in rural areas: a method and case study. *Social Science & Medicine*, 16, 85–90.
- Kalles, D., Pierrakeas, C., & Xenos, M. (2008). Intelligently Raising Academic Performance Alerts. En *Proceedings of 1st International Workshop on Combinations of Intelligent Methods and Applications* (pp. 37–42). Patras, Greece.
- Kang, J. S., Kuznetsova, P., Luca, M., & Choi, Y. (2013). Where not to eat? Improving public policy by predicting hygiene inspections using online reviews. En *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (pp. 1443–1448). Seattle, USA.
- Kankanhalli, A., Zuiderwijk, A., & Tayi, G. K. (2017). Open innovation in the public sector: A research agenda. *Government Information Quarterly*, 34(1), 84–89.

- Kleinberg, J., Ludwig, J., Mullainathan, S., & Obermeyer, Z. (2015). Prediction policy problems. *The American Economic Review*, 105(5), 491–495.
- Kokkidou, E., Rodríguez, P., & Mondaca, J. (2017). El impacto económico de la deserción escolar en Chile.
- Kovacic, Z. (2012). Predicting student success by mining enrolment data. *Research in Higher Education Journal*, 15(1), 1-20.
- Labrinidis, A., & Jagadish, H. V. (2012). Challenges and Opportunities with Big Data. *Proceedings of the VLDB Endowment*, 5(12), 2032–2033. Istanbul, Turkey
- Ledermann, R., & Madariaga, L. (2017). *Modelo de transformación digital de proyectos de modernización en el sector público del Ministerio de Hacienda*. Santiago: Ministerio de Hacienda.
- Lima, M. (2011). *Visual Complexity: Mapping Patterns of Information*. New York: Princeton Architectural Press.
- Luo, W., & Qi, Y. (2009). An enhanced two-step floating catchment area (E2SFCA) method for measuring spatial accessibility to primary care physicians. *Health & Place*, 15(4), 1100–1107.
- Luo, W., & Wang, F. (2003). Measures of spatial accessibility to health care in a GIS environment: synthesis and a case study in the Chicago region. *Environment and Planning B: Planning and Design*, 30(6), 865 – 884.
- Mansell, J., Laking, R., Matheson, B., & Light, R. (2017). The Data Commons Blueprint: A high-trust, lower-cost alternative to enable data integration and reuse. Recuperado a partir de <http://datacommons.org.nz/>
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). *Big data: The next frontier for innovation, competition, and productivity*. Washington: McKinsey & Company.
- Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). Making Sense of Data-Driven Decision Making in Education. Recuperado 19 de noviembre de 2017, a partir de https://www.rand.org/pubs/occasional_papers/OP170.html
- Mell, P., Grance, T., & others. (2011). The NIST definition of cloud computing. Recuperado a partir de <http://faculty.winthrop.edu/domanm/csci411/Handouts/NIST.pdf>
- Microsoft. (2016). Predicting student dropout risks, increasing graduation rates with cloud analytics. Recuperado 14 de diciembre de 2017, a partir de <https://customers.microsoft.com/en-us/story/tacomapublicschoolsstory>
- Microsystem. (2016). *Requerimientos de un Sistema de Gestión Documental para el Estado*. Santiago: Ministerio de Hacienda.
- Miller, H. J. (2004). Tobler's First Law and Spatial Analysis. *Annals of the Association of American Geographers*, 94(2), 284–289.
- Ministerio de Educación. (2015). *Informe de Piloto de Modelo Predictivo, Seguimiento de Estrategias de Apoyo (Sistema de Alerta Temprana)*. Santiago: Ministerio de Educación.
- Ministerio de Educación. (2017). *Ley 21.040 que crea el sistema de educación pública*. Santiago: Biblioteca del Congreso Nacional.
- Ministerio de Educación. (2017). Portal de Datos Abiertos Ministerio de Educación de Chile. Recuperado 20 de diciembre de 2017, a partir de <http://datos.mineduc.cl/home>
- Ministerio Secretaría General de la Presidencia. (1999). *Ley 19.628 sobre protección de la vida privada*. Santiago: Biblioteca del Congreso Nacional.
- Ministerio Secretaría General de la Presidencia. (2005). *Decreto 100 que fija el texto refundido, coordinado y sistematizado de la Constitución Política de la República de Chile*. Santiago: Biblioteca del Congreso Nacional.

- Ministerio Secretaría General de la Presidencia. (2008). *Ley 20.285 sobre acceso a la información pública*. Santiago: Biblioteca del Congreso Nacional.
- Ministerio Secretaría General de la Presidencia. (2013a). *Guía Rápida de Publicación de Datos Abiertos*. Santiago: SEGPRES.
- Ministerio Secretaría General de la Presidencia. (2013b). *Norma Técnica para Publicación de Datos Abiertos en Chile, versión 2.1*. Santiago: SEGPRES.
- Ministerio Secretaría General de la Presidencia. (2017). Borrador norma técnica interoperabilidad en el Estado de Chile. Recuperado a partir de http://digital.gob.cl/site/assets/files/1297/borrador_-_norma_tecnica_interoperabilidad_en_el_estado_de_chile_pdf.pdf
- Nghe, N. T., Janecek, P., & Haddawy, P. (2007). A comparative analysis of techniques for predicting academic performance. En *Frontiers In Education Conference - Global Engineering: Knowledge Without Borders, Opportunities Without Passports, 2007. FIE '07. 37th Annual* (p. T2G-7-T2G-12).
- Noticias Universidad de Chile. (2015). Commander: El software creado por la U. de Chile para acelerar la respuesta de Bomberos en casos de emergencia - Universidad de Chile. Recuperado 28 de diciembre de 2017, a partir de <http://www.uchile.cl/noticias/112527/commander-software-que-acelera-respuesta-de-bomberos-en-emergencias>
- OECD. (2010). *Overcoming school failure: policies that works*. Paris: OECD Publishing.
- OECD. (2014). *Recommendation of the Council on Digital Government Strategies*. Paris: OECD Publishing.
- OECD. (2016). *Digital Government in Chile - Strengthening the Institutional and Governance Framework*. Paris: OECD Publishing.
- OECD. (2017a). Building a data-driven public sector in Norway. En *Digital Government Review of Norway* (pp. 155-207). Paris: OECD Publishing.
- OECD. (2017b). *Government at a Glance 2017*. Paris: OECD Publishing.
- OECD. (2017c). *Innovation Skills in the Public Sector*. Paris: OECD Publishing.
- OECD. (2017d). Income Distribution Database (IDD): Gini, poverty, income, Methods and Concepts. Recuperado 19 de diciembre de 2017, a partir de <http://www.oecd.org/social/income-distribution-database.htm>
- Open Knowledge International. (2017). Global Open Data Index. Recuperado 30 de diciembre de 2017, a partir de <https://index.okfn.org/>
- Open Knowledge International. (2017). What is open data?. Recuperado 20 de diciembre de 2017, a partir de <http://opendatahandbook.org/guide/es/what-is-open-data/>
- OpenStreetMap. (2017). Recuperado 21 de diciembre de 2017, a partir de <http://www.openstreetmap.org/>
- Pal, S. (2012). Mining Educational Data Using Classification to Decrease Dropout Rate of Students. *International journal of multidisciplinary sciences and engineering*, 3(5), 35-39.
- Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data*, 1(1), 51-59.
- Radio Villa Francia, E. (2016). Sename reconoce en su informe la escalofriante cifra de 1.313 muertes bajo su tutela en un periodo de 11 años. Recuperado 19 de diciembre de 2017, a partir de <http://www.radiovillafrancia.cl/sename-reconoce-en-su-informe-la-escalofriante-cifra-de-1-313-muertes-bajo-su-tutela-en-un-periodo-de-11-anos>
- Radke, J., & Mu, L. (2000). Spatial Decompositions, Modeling and Mapping Service Regions to Predict Access to Social Programs. *Geographic Information Sciences*, 6(2), 105-112.

- Rodríguez, P. (2017,). La Geografía de la Educación: Respondiendo preguntas estratégicas para la instalación y gestión de los Servicios Locales de Educación Pública.
- Rodríguez, P., Palomino, N., & Mondaca, J. (2017). *El uso de datos masivos y sus técnicas analíticas para el diseño e implementación de políticas públicas en Latinoamérica y el Caribe* (Discussion Papers & Presentations No. IDB-DP-514). Washington: Inter-American Development Bank.
- Rodríguez, P., & Pinto, C. (2017). Memoria de cálculo de impacto económico de la deserción y repitencia por subvenciones perdidas.
- Rouse, C. E. (2005). The Labor Market Consequences of an Inadequate Education. In *Symposium on the Social Costs of Inadequate Education, Teachers College Columbia University*. New York: USA.
- Schmarzo, B. (2013). *Big Data: Understanding How Data Powers Big Business*. Indianapolis: John Wiley & Sons.
- Song, I.-Y., & Zhu, Y. (2016). Big Data and data science: what should we teach? *Expert Systems*, 33(4), 364–373.
- Staff Science. (2011). Dealing with Data: Challenges and Opportunities. *Science*, 331(6018), 692–693.
- Steele, J., & Illinsky, N. (Eds.). (2010). *Beautiful Visualization: Looking at Data through the Eyes of Experts* (1ª ed.). Sebastopol: O'Reilly Media.
- Thai-Nghe, N., Horváth, T., & Schmidt-Thieme, L. (2011). Factorization Models for Forecasting Student Performance. En *Proceedings of the 4th International Conference on Educational Data Mining* (pp. 11–20). Eindhoven, The Netherlands.
- The Economist. (2010). Data, data everywhere. *The Economist*. Recuperado a partir de <http://www.economist.com/node/15557443>
- The World Wide Web Foundation. (2016). *Open Data Barometer 3rd Edition*. Washington, D.C: The World Wide Web Foundation.
- The World Wide Web Foundation. (2017). *Open Data Barometer 4th edition*. Washington, D.C.: The World Wide Web Foundation.
- Tomar, L., Guicheney, W., Kyarisiima, H., & Zimani, T. (2016). *Big Data in the Public Sector: Selected Applications and Lessons Learned* (Discussion Papers & Presentations). Washington D.C.: Inter-American Development Bank.
- Ubaldi, B. (2013). *Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives*. Paris: OECD Publishing.
- UNESCO. (2012). *Policy brief - Learning Analytics*. New York: UNESCO Institute for Information Technologies in Education.
- Wei, Z. (2013). *A study of accessibility to health facilities for elderly population in metro Atlanta using a categorical multi-step floating catchment area method* (Thesis). uga. Recuperado a partir de <http://ugakr.libs.uga.edu/xmlui/handle/10724/28942>
- Wiseman, J. (2015). *Innovations in Public Service Delivery: Issue No 01: Can 311 Call Centers Improve Service Delivery? Lessons from New York and Chicago* (Discussion Papers & Presentations). Washington D.C.: Inter-American Development Bank.
- Wiseman, J. (2016). *Analytics Excellence Roadmap*. Cambridge: Harvard Kennedy School, Ash Center for Democratic Governance and Innovation.
- Wiseman, J. (2017a). *Innovations in Public Service Delivery: Issue No. 5: Improving Service Delivery through Information Integration: Building a Single View of the Citizen* (Discussion Papers & Presentations). Washington D.C.: Inter-American Development Bank.

- Wiseman, J. (2017b). Discovering the True Value of City Data Experts. Recuperado 30 de diciembre de 2017, a partir de <https://datasmart.ash.harvard.edu/news/article/discovering-the-true-value-of-data-experts-who-are-showing-cities-the-money>
- World Bank Group. (2016). *Big Data Innovation Challenge*. New York: World Bank.
- World Bank Group. (2017). GDP per capita, PPP (current international \$) | Data. Recuperado 19 de diciembre de 2017, a partir de <https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD?locations=CL>
- Yau, N. (2011). *Visualize This: The FlowingData Guide to Design, Visualization, and Statistics* (1ª ed.). Indianapolis: John Wiley & Sons.

ANEXO: EJEMPLOS DE METODOLOGÍAS ASOCIADAS A LA GESTIÓN DE DATOS Y SERVICIOS TECNOLÓGICOS AFINES

A modo de complementar las metodologías y procesos de analítica desarrollados en la sección 1.3, este anexo explica algunas técnicas de procesamiento y servicios tecnológicos para el procesamiento de datos masivos. Estas definiciones son importantes, para distinguir entre técnicas analíticas y productos específicos que las implementan, en particular para el caso de los servicios tecnológicos. Algunas metodologías relevantes asociadas al procesamiento de datos masivos son descritas en la Tabla A1, mientras que la Tabla A2 expone algunos ejemplos relevantes de servicios tecnológicos.

Tabla A1: Metodologías asociadas a la gestión de datos para el procesamiento de ‘Big Data’ .(Rodríguez et al., 2017)

Tipo de datos	Ejemplos de técnicas de procesamiento según el tipo de datos
Texto	Extracción de la información: obtener datos estructurados de un texto reconociendo entidades y relaciones entre ellas. Resumen de texto: construir resúmenes de uno o múltiples documentos usando procesamiento de lenguaje natural. Respuesta a la pregunta: responde a preguntas formuladas en lenguaje natural usando procesamiento desarrollado para tales fines. Análisis de sentimiento: analiza un texto de opinión y genera una respuesta negativa o positiva.
Audio	Enfoque basado en transcripción: se genera una transcripción textual de contenido del audio a través de reconocimiento automático del habla con grandes diccionarios. El resultado se analiza con las técnicas analíticas de texto. Enfoque basado en la fonética: trabaja a partir de sonidos o fonemas que se traducen a una secuencia a partir del habla. La representación fonética de un término se busca en la secuencia.
Video	Arquitectura basada en el servidor: servidor dedicado al análisis de videos. Arquitectura basada en el borde: el video es analizado en forma local y sobre el video sin compresión de datos.
Redes sociales	Analítica basada en contenido: se enfoca en los datos posteados por los usuarios, que son analizados posteriormente con técnicas descritas anteriormente, sea texto, audio o video. Analítica basada en la estructura: sintetiza los atributos estructurales de la red social y extrae inteligencia a partir de las relaciones de las entidades participantes. Las técnicas incluyen detección de comunidades, análisis de influencia social y predicción de enlaces.

Tabla A2: Ejemplos de servicios tecnológicos utilizados para manipular y analizar grandes volúmenes de datos .(Rodríguez et al., 2017).

Servicio	Descripción	Ejemplos
Nuevos frameworks analíticos	Son entornos de trabajo que contienen o pueden contener una serie de paquetes y librerías que permiten reutilizar código, facilitando tareas comunes.	Hadoop (Google, Apache), Spark.
Almacén de datos y Lagos de datos (Data warehouse & Data lakes)	Son repositorios de datos a los que sólo se insertan nuevos datos y, a través de los llamados <i>data marts</i> , permiten la generación de bases de datos. La diferencia entre <i>data warehouse</i> y <i>data lake</i> radica en la estructura de los datos. Los <i>data warehouses</i> almacenan datos estructurados (vale decir, tablas con datos ordenados en filas y columnas), mientras que los <i>data lakes</i> pueden almacenar datos estructurados, semi-estructurados e inestructurados	SQL Server, Azure SQL, NoSQL.
Base de datos relacional (relational database)	Base de datos compuesta de una colección de tablas (relaciones). Los sistemas de gestión de bases de datos relacionales (RDBMS) almacenan un tipo de datos estructurados. SQL es el lenguaje más utilizado para la gestión de bases de datos relacionales (véase el tema más adelante).	MySQL, PostgreSQL, Oracle, SparkSQL.
Base de datos no relacional (non-relational database)	Base de datos que no almacena los datos en formato tabular (filas y columnas).	MongoDB, Cassandra
Visualización de datos	Herramientas que permiten la visualización de datos. Se diferencian comúnmente en la flexibilidad y versatilidad que ofrecen para crear visualizaciones personalizadas.	D3js, Google Charts, Tableau, Vega

Servicio	Descripción	Ejemplos
Herramientas/ <i>Plugins</i> estadísticos	Refiere a paquetes estadísticos o extensiones, que permitan realizar análisis estadísticos sobre los datos. Los más sencillos usualmente permiten realizar todas las tareas a través de una interfaz gráfica, mientras que los más complejos requieren conocimientos de algún lenguaje de programación.	SAS, Stata, SPSS, Matlab, R, Python Pandas
Sistema de información geográfica (<i>Geographic information system</i>)	Son sistemas diseñados para manipular, almacenar, analizar y visualizar datos geográficos.	Leaflet, PostGIS, Esri ArcGIS, CartoDB
Servicios en la nube (<i>Cloud Computing</i>)	Es un modelo que permite acceso <i>on-demand</i> a una serie de recursos computacionales configurables. Estos recursos pueden ser utilizados o liberados sin necesidad de interacción con el proveedor y con pocos recursos a nivel de gestión (Mell, Grance, & others, 2011). Cinco se reconocen como las características esenciales del Cloud Computing: auto servicio <i>on-demand</i> , mayor acceso a red (<i>broad network access</i>), conjunto de recursos (<i>resource pooling</i>), rápida elasticidad y capacidad para controlar y optimizar el uso de recursos de forma automática.	Amazon Web Services, Google Cloud Platform, Microsoft Azure, Digital Ocean: Cloud Services for Developers.