

DOCUMENTO
DE TRABAJO

Nº **16**

Datos abiertos para la investigación
en educación: Hacia el desarrollo
de estándares para el registro y
almacenamiento de datos en el
Sistema de Aseguramiento de la
Calidad (SAC)

María Victoria Martínez M., Hadabell
Castillo H., Cristian Grandón G.

Septiembre, 2018

DOCUMENTO
DE TRABAJO N ° 16

DATOS ABIERTOS ARA LA INVESTIGACIÓN EN EDUCACIÓN:
HACIA EL DESARROLLO DE ESTÁNDARES PARA EL REGISTRO Y
ALMACENAMIENTO DE DATOS EN EL SISTEMA DE
ASEGURAMIENTO DE LA CALIDAD (SAC)

María Victoria Martínez M., Hadabell Castillo H., Cristian
Grandón G.
Septiembre, 2018

Datos abiertos para la investigación en educación: Hacia el desarrollo de estándares para el registro y almacenamiento de datos en el Sistema de Aseguramiento de la Calidad (SAC)

Centro de Estudios MINEDUC
División de Planificación y Presupuesto
Subsecretaría de Educación
Ministerio de Educación, República de Chile
Av. Libertador Bernardo O'Higgins N° 1371 Santiago, RM
Tel. 22 406 6000
©2018 Ministerio de Educación

Presidente de la República de Chile:
Sebastián Piñera E.

Ministra de Educación:
Marcela Cubillos S.

Subsecretario de Educación:
Raúl Figueroa S.

Subsecretaria de Educación Parvularia:
María José Castro R.

Jefe de División de Planificación y Presupuesto, Subsecretaría de Educación:
Leon Paul C.

Jefe de Centro de Estudios (S), Subsecretaría de Educación:
Bárbara Manríquez G.

Autora publicación:
María Victoria Martínez M., Hadabell Castillo H., Cristian Grandón G.

Coordinación general de la publicación:
Unidad de Promoción y Difusión de la Investigación

Edición y diagramación:
Daniela Ubilla R.

En la presente publicación se aplican las Orientaciones para un uso de lenguaje no sexista e inclusivo elaborado por la Unidad de Equidad de Género del Ministerio de Educación.

Se autoriza su reproducción siempre y cuando se haga referencia explícita a la fuente.

Para referenciar, emplear el siguiente formato:

Ministerio de Educación, Centro de Estudios (2018). Datos abiertos para la investigación en educación: Hacia el desarrollo de estándares para el registro y almacenamiento de datos en el Sistema de Aseguramiento de la Calidad (SAC). *Documento de trabajo N° 16*. Santiago, Chile.

Índice

Resumen	5
1. Introducción.....	7
2. Datos abiertos: Definiciones, implicancias y limitaciones para el uso en investigación.....	9
3. Desarrollo histórico de la implementación de datos abiertos.....	15
4. Problema de investigación.....	21
5. Metodología.....	21
6. Resultados.....	23
7. Conclusiones.....	28
8. Orientaciones y formatos para registro y almacenamiento de datos cuantitativos y cualitativos en el marco del Sistema de Aseguramiento de la Calidad de la Educación (SAC).....	31
Bibliografía.....	39

Resumen

Con la finalidad de dar a conocer los fundamentos, las condiciones y los estándares necesarios que permitan potenciar la utilización y reutilización de los datos generados a partir de investigaciones en educación, este documento expone un marco conceptual y una revisión de antecedentes que enmarca los fundamentos teóricos y prácticos tras el enfoque de datos abiertos.

Posteriormente se presentan los resultados de la revisión realizada a los sitios web de 23 ministerios de la República de Chile cuyo objetivo fue conocer cómo se realiza en organismos públicos la publicación de información con fines investigativos.

Los resultados se encuentran agrupados en tres dimensiones: disponibilidad, acceso y reutilización. En tanto disponibilidad se constata que mayoritariamente en los sitios web hay informes de investigación o bases de datos y, de manera cada vez más emergente, se encuentran disponibles datos georreferenciados. No fue posible encontrar datos de índole cualitativo. En tanto acceso, se identificó que la utilización de buscadores especializados de información no es un recurso muy empleado por las instituciones y que son escasos los organismos que entregan datos en distintos tipos de formato. En tanto reutilización, los resultados indican que no es una práctica institucionalizada ni generalizada hacer disponibles otros materiales asociados a las investigaciones (instrumentos, transcripciones o bases de datos) Exceptuando algunos casos, tampoco se estandariza la entrega de la información ni tampoco se incluyen variables que permitan la interoperabilidad de los datos.

A partir de la revisión de la literatura y de la revisión de los casos empíricos, se concluye que la apertura de datos es un elemento democratizador que permite la generación de evidencia para la política pública desde distintos sectores de la ciudadanía, pero que sin embargo, actualmente a nivel gubernamental coexisten experiencias que se encuentran en diferentes niveles de progreso, es decir, existen distintos estadios en la implementación de datos abiertos. A partir de esto se identifica una oportunidad para el Sistema de Aseguramiento de la Calidad de mejorar sus estándares de generación de información que permita una reutilización de los datos que genera.

Con la reutilización de los datos se identifica que, por un lado, aumenta la posibilidad de disminuir la cantidad de procesos de recolección de información en los establecimientos escolares del país, y de reducir costos económicos y de tiempo tanto de quienes investigan como de las distintas comunidades que participan en estos procesos. Y por otro, permite avanzar en la implementación de un gobierno digital que utilice más, de mejor manera y por medio de la interacción y colaboración con la ciudadanía, evidencia para la toma de decisiones de las políticas educativas que aspiran mejorar la calidad de vida de las personas.

Para finalizar, el documento presenta estándares para guiar el desarrollo de investigaciones en educación en el Sistema de Aseguramiento de la Calidad que permitirán la reutilización de los datos que se generan a partir de investigaciones.

Abstract

Aiming to announce the basics, conditions and standards necessary to enhance the use and reuse of data generated from research in education, this document presents a conceptual framework and a background check that frames the theoretical and practical foundations under the open data approach.

Later, the results of the review made to the websites of 23 ministries of the Republic of Chile are presented, whose objective was to know how the publication of information for investigative purposes is carried out in public organisms.

The results are grouped into three dimensions: availability, access and reuse. Regarding availability, it is found that in the web sites there are mostly research reports or databases and, in an increasingly emergent way, georeferenced data are available. It was not possible to find qualitative data. In terms of access, it was identified that the use of specialized information search engines is not a resource widely used by institutions and that there are few agencies that provide data in different types of format. Regarding reuse, the results indicate that it is not an institutionalized or generalized practice to make available other materials associated with research (instruments, transcriptions or databases). Except for some cases, the delivery of information and the inclusion of variables that allow interoperability of the data is not standardized.

The review of the literature and the review of empirical cases, allows us to conclude that the opening of data is a democratizing element that allows the generation of evidence for public policy from different sectors of the citizenry, however, currently at the governmental level experiences that are at different levels of progress coexist, that is, there are different stages in the implementation of open data. This allows us to identify an opportunity to improve the standards of information generation that allows a reuse of the data generated by the Sistema de Aseguramiento de la Calidad (Quality Assurance System).

With the reuse of the data, it is identified that it increases the possibility of reducing the amount of information gathering processes in schools in the country, and the reduction of economic costs as well as the time of both those who investigate and the different communities that participate. It also allows progress in the implementation of a digital government that uses through interaction and collaboration with citizens, more and better, evidence for decision making of educational policies that aim to improve the quality of life of people.

Finally, the document presents standards to guide the development of research in education in the Sistema de Aseguramiento de la Calidad that will allow the reuse of data generated from research.

1. Introducción

Durante los últimos años ha comenzado una revolución de la información a nivel mundial que tiene directa relación con el acceso libre a los datos. Open Data (Datos Abiertos) responde a una iniciativa ya instalada en diversos países y que persigue la publicación de conjuntos de datos que organizaciones, empresas, instituciones, el Estado, entre otros, tienen en su poder, con el fin de que estos sean divulgados y potencialmente reutilizados por la sociedad en general, y, en particular, por entidades que puedan agregar valor a dichos datos y así elaborar productos beneficiosos para la comunidad (Minsegespres, 2013, p. 3). Esta transformación en que lo privado pasa a tener un acceso libre, atañe inicialmente al sector privado (negocios y empresas) para luego tener consecuencias en el aumento de expectativas en el sector público; así, se entrega información útil y reutilizable con un propósito no reactivo, sino que proactivo y se genera una relación colaborativa entre el Estado y la ciudadanía. De esta forma, esta revolución ha provocado que el Estado no se encargue solamente de generar información y aplicaciones para la ciudadanía, sino que sea esta misma la que potencialmente saque provecho de los datos que son puestos a disposición (Concha & Naser, 2012).

En este panorama de transformación a nivel internacional, ocurre también un giro a nivel nacional en relación con la apertura de la información; así, en el año 2005 se generó en Chile una moción parlamentaria que buscó la tramitación del proyecto de ley sobre acceso a la información pública. Al respecto, aunque diversas legislaciones del mundo han realizado esfuerzos con el objetivo de consagrar el derecho a la transparencia de la información, y que Chile también ha intentado en los últimos años, aun es necesario suplir las deficiencias prácticas de la legislación sobre el acceso a la información pública (Larraín & Gazmuri, 2005). En consecuencia, se promulgó en Chile la ley sobre acceso a la información pública el año 2008, la que buscó respetar y cautelar la publicidad de los actos, resoluciones, procedimientos y documentos de la administración, así como la de sus fundamentos, para facilitar el acceso de cualquier persona a esa información mediante los medios y procedimientos que al efecto establezca la ley.

La promulgación de la ley de acceso a la información pública, también llamada “Ley de Transparencia”, instauró una responsabilidad en todas las instituciones públicas, incluyendo los 23 ministerios como órganos superiores de colaboración de la administración del Estado chileno. Dentro de estos, el Ministerio de Educación (en adelante MINEDUC) es el ente responsable (entre otras funciones) de desarrollar estadísticas, indicadores y estudios del sistema educativo en el ámbito de su competencia, y poner a disposición del público la información que con motivo del ejercicio de sus funciones recopile. Esta información será pública y de libre acceso para quien tenga interés en consultarla (Ley N° 20.529, Artículo N° 111, p. 40).

Dentro del MINEDUC, el Centro de Estudios cumple la función de contribuir a los propósitos y objetivos de este Ministerio mediante la producción, promoción y difusión de información y conocimiento con altos estándares de calidad. Se busca generar material pertinente al desarrollo

de las políticas públicas y a los contenidos disponibles, que sea de acceso público y que está dirigido tanto a quienes se dedican a la investigación como a la comunidad educativa.

Al considerar la labor del Centro de Estudios en la generación, promoción, articulación y difusión de datos e investigación, es que durante los últimos años se ha impulsado una política activa de difusión que promueve el principio de transparencia activa planteado en la ley. Por un lado, por medio del Centro de Documentación (CEDOC) que posee desde el año 1999 a la fecha más de 600 documentos correspondientes principalmente a informes, reportes y artículos que dan cuenta de los distintos procesos investigativos que se han realizado en o para el Ministerio de Educación. Por otro, mediante el portal de Datos Abiertos MINEDUC, el cual es una plataforma de vanguardia en el contexto latinoamericano, con el mayor acceso a datos desagregados en educación entre los países de la OCDE y que posee más de 350 bases de datos procedentes de registros administrativos (Centro de Estudios, 2018).

No obstante esta gran cantidad de información disponible, y buscando potenciar aún más el libre acceso a la información generada con recursos públicos, el Centro de Estudios se ha propuesto ampliar el espectro de información disponible en el portal de Datos Abiertos MINEDUC, incorporando bases de datos de otras instituciones pertenecientes al Sistema de Aseguramiento de la Calidad y también datos generados a partir del desarrollo y ejecución de investigaciones elaboradas o financiadas por el MINEDUC.

Este desafío requiere identificar cuáles son los estándares necesarios para el registro y almacenamiento de datos e información de calidad que permitan generar orientaciones para el Sistema de Aseguramiento de la Calidad en Educación.

A partir de lo dicho, este documento de trabajo expone un marco conceptual y una revisión de antecedentes que enmarca los fundamentos teóricos y prácticos tras el enfoque de datos abiertos; posteriormente, se presentan los resultados de la revisión de sitios web nacionales que permiten, por un lado, conocer cómo distintas instituciones abordan el libre acceso a los datos y, por otro, identificar iniciativas destacadas en esta línea; se finaliza con la presentación de estándares que guiarán el desarrollo de investigaciones en educación en el Sistema de Aseguramiento de la Calidad (SAC).

2. Datos abiertos: definiciones, implicancias y limitaciones para el uso en investigación

La presente sección expone los conceptos básicos asociados a datos abiertos, esto es, los conceptos relacionados con el contexto actual del uso y disponibilidad de datos abiertos, y los criterios desde los cuales es posible hablar de ellos. Asimismo, serán abordados los datos abiertos en contexto de gobierno u *open government*.

Luego de ser expuestas las oportunidades en la generación de datos abiertos serán presentadas las potenciales amenazas, retos y barreras en el uso que enfrentan investigadores e investigadoras en la generación, uso y reutilización de esta información. Posteriormente, se revisarán las limitaciones respecto de su uso en investigación.

Definición e implicancias de los datos abiertos

El contexto actual da cuenta de una sociedad que desarrolla y acumula un sinnúmero de información, pero que presenta, a su vez, el desafío de almacenarla y mostrarla de una manera válida y pertinente para que sea accesible y esté disponible para su uso. Sin embargo, aunque se ha establecido la necesidad de compartir información, actualmente se busca permitir el libre acceso no solo a publicaciones, sino que también a los datos, poniéndolos a disposición de manera libre y gratuita (Ferrer-Sapena & Sánchez-Pérez, 2012). A este respecto, la comisión europea de 2014 consideró que los datos derivados de la investigación y los resultados de experimentos científicos eran especialmente valiosos para la economía de los países (Unión Europea, 2014).

De esta manera, se puede afirmar que por la cantidad de datos y por su disponibilidad, se está propiciando un verdadero cambio de paradigma que se basa en lo abierto (Peset, F. et al., 2017), donde la explotación efectiva de datos masivos depende fundamentalmente de una cultura internacional de datos abiertos (Boulton, 2015) en que se hace necesario integrar datos procedentes de un gran número de fuentes de manera dinámica y actualizada (Peset, F. et al., 2017).

Se entiende como open data o datos abiertos la apertura o puesta a disposición de datos digitales a cualquier persona o institución. La promoción del acceso entonces está vinculada a la transparencia de los procesos y, dependiendo de quién es el productor de los datos, se hablará de transparencia en el gobierno, en la investigación o en las empresas (Ferrer-Sapena & Sánchez-Pérez, 2012). A este respecto, la Open Knowledge Foundation (2012) especifica que los datos abiertos de cualquier origen deben poder ser: utilizados (deben estar disponibles en internet y permitir su modificación); reutilizados y redistribuidos (en condiciones óptimas para combinarse con otros conjuntos de datos); y de acceso libre a cualquier persona (donde todos son capaces de utilizarlos, reutilizarlos y redistribuirlos sin restricción), permitiendo la creación de obras derivadas que reconozcan la misma licencia. Esto significa que hacerlos disponibles o liberarlos al público no los incluye automáticamente en la órbita de lo abierto (Peset, F. et al., 2017), es necesario que

cumplan con los Principios FAIR (Doorn & Dillo, 2016; Méndez, 2016) esto es que sean accesibles (guardados por un tiempo prolongado, siendo fácil de acceder y descargar), localizables (fáciles de encontrar), interoperables (posible de ser combinado/articulado con otros datos) y reutilizables (en futuras investigaciones o usando diferentes métodos) . Por esta razón, la Royal Society (2012) definió los datos abiertos como aquellos que cumplen ciertos criterios cuando son liberados, lo que se llama apertura inteligente/competitiva.

Debido a que esta disponibilidad de datos no es inherente al proceso de generación de información sino que responde a la tendencia actual de apertura, es que se hace necesario identificar las formas en que actualmente se hacen públicos los datos. Si bien, ya se ha superado la privacidad completa de los datos se identifica en la actualidad un proceso de transición desde los primeros elementos de apertura de datos hacia el modelo ideal de apertura inteligente y competitiva. A este respecto, Peset, F. et al. (2017) presentan las dos formas básicas en que hoy se hacen públicos los datos: a) La vía verde del acceso abierto, que se refiere a la generación de repositorios de información vinculados a la comunidad científica, a gestores de información y empresas y, finalmente, a las instituciones. Esta vía ha sido explorada y se asocia a garantizar la preservación de la información y a respetar la interoperabilidad; usualmente es empleada en círculos de investigación o instituciones específicas sin tener mayor apertura a personas externas al ámbito en cuestión. Por lo general, esta vía es de acceso libre de pago y se comporta como un repositorio de documentos, los cuales no necesariamente son revisados por pares o se presentan en su última versión (Datos abiertos Gobierno de España, 2017); b) La vía dorada del acceso abierto supone el almacenamiento de material suplementario además del material central, como un artículo o la publicación de data papers, dando espacio a nuevos modelos de publicación donde se difunden además de las investigaciones, los materiales referentes a resultados, documentos y materiales suplementarios a ellas (Peset, F. et al., 2017). Respecto este segundo modelo, este se relaciona usualmente con un sistema pagado de acceso a la información, por lo que su acceso es de mayor exclusividad.

Si bien la vía dorada se percibe como la solución que se acerca más a los principios FAIR no los cumple completamente, lo que releva elementos necesarios de seguir desarrollando; ejemplo de ello es la necesidad de proporcionar más material suplementario a los artículos, la falta de interoperabilidad entre los datos de los almacenamientos existentes, la escasa garantía de preservar los datos en el futuro, y la poca claridad en las políticas sobre lo que se puede hacer o no con la información, y la necesidad de eliminar el acceso restringido a suscriptores (Peset, F. et al., 2017). En este contexto, cada sistema productor de información deberá propender a una apertura inteligente cercana a la vía dorada.

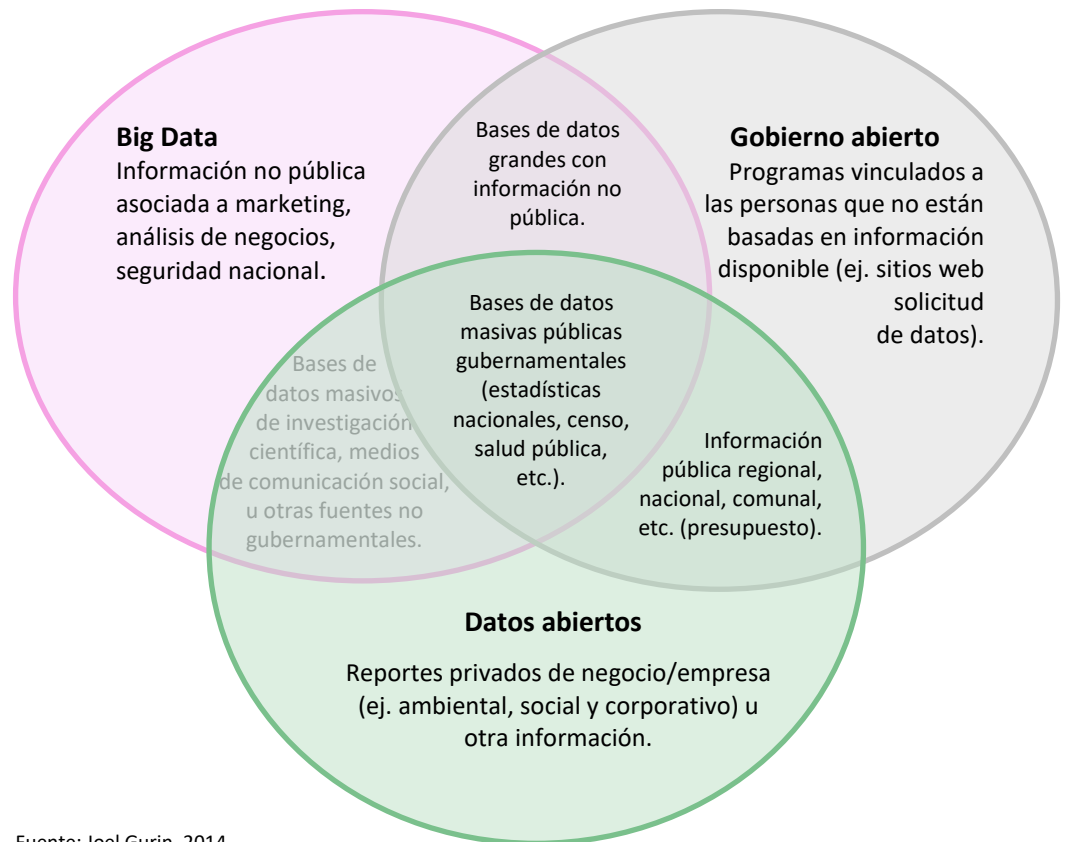
Por otra parte, existe también una reflexión respecto de los deseos y expectativas de quienes publican los datos en relación con la reutilización, y los propósitos de uso de los elementos presentados. González & Peset (2013) señalan la necesidad de que, al momento de difundir los datos, la publicación se debe acompañar de una declaración manifiesta y explícita que detalle cuáles son las expectativas.

La mayoría de las administraciones expone cuáles son sus políticas de reutilización, dada la multitud de aplicaciones y servicios que se pueden generar. En Reino Unido, por ejemplo, indican en qué casos pueden o no utilizarse. Como es lógico, prohíbe su uso para actividades ilegales, violentas, racistas, cambiar o tergiversar los datos, etc. Se puede publicar, distribuir, adaptar la información y explotarla comercialmente si se combina con el uso de otros datos. Las empresas deben citar la fuente de los datos y, a ser posible, hacer un enlace a la licencia de uso (Ferrer-Sapena et al., 2011).

Una especificación del uso de la información disponible es necesaria entonces tanto para resguardar la validez en el uso de la información publicada como para asegurar su interoperabilidad y el uso futuro de esta, en otras palabras, la necesidad base de resguardar la privacidad y el uso adecuado de información de los individuos o instituciones de los que fue obtenida la información.

Según las características que deben tener los datos puestos a disposición y que ya fueron presentadas, es preciso comprender desde qué contexto se genera el acceso a los datos. En términos concretos, si bien la promoción del acceso a la información depende de las características de la información que se hace disponible, también es determinada por el sistema que la produce. Este contexto queda representado en la siguiente figura que simboliza el ecosistema de disponibilización de datos conformado por los sistemas de datos masivos o big data, datos abiertos u open data y, por gobierno abierto u open government. De manera gráfica, se muestran tres sistemas que al combinarse conforman un ecosistema de información, y que, por tanto, tienen diferentes implicancias en cuanto a la apertura de la información. Por ejemplo, gobierno abierto refiere a la información que se disponibiliza a la ciudadanía de manera reactiva, es decir, luego de una solicitud de información. Sin embargo, en la medida que el sistema de gobierno abierto se complementa con el sistema de datos abiertos se logra un panorama de datos abiertos de gobierno, en el cual es posible generar una apertura inteligente de los datos que ponen a disposición diferentes materiales de acceso libre para ser utilizados y reutilizados por cualquier persona, siendo entonces un ecosistema proactivo de información en que ciudadanía y gobierno trabajan colaborativamente.

Esquema 1.



Fuente: Joel Gurin, 2014.

Un sistema inteligente de información entonces se comprende por tres sistemas principales aunque, por ejemplo, se entiende que el open government (gobierno abierto) forma parte, ciertamente no es la única forma en que este se comprende. Se observa entonces que es una de las tres partes del ecosistema de información masiva que existe a disposición (Back, B. 2016), siendo las otras partes la big data y la open data. Respecto a estas últimas, mientras que la open data refiere a la información disponible para el uso público sin importar el propósito de este, la big data refiere a bases de datos de gran escala (masivas) y que, por lo tanto, no pueden ser empleadas sin el manejo de un software especializado. Así, mientras que los datos abiertos son definidos por su uso, el big data es definido por su tamaño. De esta manera, en la medida en que estos tres sistemas se complementan se logra un ecosistema inteligente que cumple con las características de apertura de la información cuando los datos son liberados, es decir, son de libre acceso, se encuentran disponibles a ser utilizados y son potencialmente reutilizados (Open Knowledge Foundation, 2012).

Respecto de la aplicación de la open data en un modelo de gobierno, si bien generar un sistema abierto puede tener implicancias políticas, al abrir la discusión sobre diferentes aspectos que no son inicialmente de conocimiento público, el acceso a la información puede aportar

razonamientos sobre porqué se han tomado determinadas decisiones, e incluir argumentos respecto de las razones para haberlas apoyado (Ferrer-Sapena & Sánchez-Pérez, 2012). En complemento, y en cuanto a lo que se refiere a la información del sector público, Ferrer-Sapena (2017) menciona que se entiende por información del sector público (ISP) todo documento, cualquiera que sea su soporte material o electrónico, así como su forma de expresión, gráfica, sonora o en imagen, que haya sido elaborado o custodiado por cualquiera de los órganos u organismos de la Administración Pública (Ferrer-Sapena, 2017, p.26). Asimismo, Noveck (2012) entiende que un sistema de datos abiertos aporta a una democracia compleja y a un proceso de innovación en un gobierno, donde se logra un gobierno más transparente y responsable, que posibilita la evidencia de malas prácticas y el rediseño de las instituciones, facilita la creación de soluciones colaborativas, y posibilita la participación de terceros, generando la transformación de la institución (Ferrer-Sapena & Sánchez-Pérez, 2012). De esta manera, se ha desarrollado también el concepto de open government o gobierno abierto que refiere a la comunicación abierta, permanente y bidireccional entre la administración y la ciudadanía, que tiene como finalidad conseguir una colaboración real entre todos y todas (Ferrer-Sapena, et al., 2011).

A su vez, la OCDE entiende, dentro del mismo contexto de apertura a la información, que la transformación del sector privado en relación con la disponibilidad y uso de la información ha generado nuevas expectativas del sector público. Por consiguiente, aunque históricamente el sistema público, en la forma de gobierno, ha tendido a entregar la información con un fin reactivo en respuesta a la ciudadanía, en el contexto actual de datos abiertos se desarrolla un giro al propósito entendiéndolo como un propósito proactivo donde el gobierno y la ciudadanía participan de un mismo proceso de colaboración y decisión. Esta transformación tiene como consecuencia el paso de un gobierno electrónico que hace disponible alguna información, a un gobierno digital, donde el uso de tecnologías digitales es asumido como parte integral de los procesos de modernización e innovación de los gobiernos, y que ofrece una mejor respuesta a las demandas (OCDE, 2017) logrando cambiar el giro hacia un gobierno que es impulsado por la ciudadanía con un sistema más interactivo y colaborativo.

Limitaciones de datos abiertos en investigación

Debido a la gran cantidad de información disponible y al creciente acceso de esta, es necesario reflexionar en las normas éticas que implican la recolección, gestión y uso que hacen las organizaciones de los datos (Datos Gobierno, 2017). Si bien es posible pensar que cada organización es responsable de establecer sus propias pautas éticas, existen también pautas desarrolladas por organismos internacionales como la Comisión Europea, la Research Data Alliance o el Gobierno del Reino Unido (Datos Gobierno, 2017) que marcan el hilo conductor de los países que comienzan a avanzar en este ámbito.

De esta forma, aunque existe consenso en considerar que la publicación de la información con acceso abierto y la posibilidad de compartir y reutilizar los datos ofrece una serie de oportunidades para el progreso del conocimiento y la generación de beneficios económicos, también se asume la necesidad de considerar que este libre acceso trae aparejado un conjunto de

amenazas y retos para los equipos de investigación (Manny, 2017). En este sentido, la ética de los datos se relaciona con las decisiones, prácticas y procesos que se realizan en una investigación y que tiene el potencial de tener un impacto en las personas, tanto directa como indirectamente (Open Data Institute, 2017).

Al respecto, el Reglamento General de Protección de Datos de la UE¹ (2016) estipula que cada individuo tiene el derecho de conocer de qué manera se están manipulando sus datos personales, con el objeto de entregar mayor control de la información personal en un contexto expansivo de magnitud de la información (Open Data Institute, 2017). Los principios de este reglamento de protección responden a que: a) el uso de la información debe tener beneficios claros para quienes participan y servir al bien común; b) la identidad de quienes participan (individuos o instituciones) está protegida y la información recogida es confidencial y segura; c) los métodos utilizados son consistentes con estándares reconocidos de integridad y calidad; d) la información y metodología empleada es consistente con los requerimientos legales como derechos humanos, reglamentos de protección y legislación en general; e) las opiniones del público se consideran a la luz de los datos manejados y los beneficios percibidos por la investigación; y f) el acceso, uso e intercambio de datos es transparente y se comunica de forma clara y accesible al público (UK Statistics Authority, 2018).

Asimismo, el European Data Portal (2016) menciona que existen diversas formas de manejar el riesgo de publicar información personal. De esta manera, un programa exitoso y sostenible de datos abiertos debe estar basado en tres pilares: i) el moral, que refiere a que quien publica la información debe considerar la privacidad de los datos, ii) el legal, donde la protección de la información debe ser respetada, y iii) el pragmático, refiriéndose a que la confianza pública debe mantenerse.

En este sentido, y debido a que la realización de análisis con gran cantidad de datos se ha expandido tanto en los sectores públicos como privados, comienzan a existir implicancias en cuanto a la privacidad, la protección de la información y los derechos asociados a las personas de las que se ha obtenido la información y de las cuales las instituciones deben hacerse cargo. Se crean entonces nuevas oportunidades, pero también muchos desafíos que antes no eran considerados. Así, las organizaciones, además de tener que ser transparentes en cuanto a la información que manejan, tendrán que ser responsables del uso que hacen de la información personal de los individuos. Sin embargo, a pesar de considerar esto se estipula que el hecho de que existan implicancias no genera necesariamente una barrera, ya que cuidar la privacidad no es un fin en sí mismo sino un derecho necesario de resguardar (Information Commissioner's Office, 2017).

Por otra parte, existen también ciertas limitaciones que refieren a barreras de uso de los datos abiertos. En Chile, el Ministerio Secretaría General de la República (2016) los expone como principales problemas en la implementación de la iniciativa de datos abiertos: barreras culturales

¹ European General Data Protection Regulation, GDPR.

(aprensión al mal uso de la información), escasa cultura de reutilización de datos, desconocimiento de qué es un dato abierto, desconfianza al mal uso de los datos por terceros, venta de información pública por parte de algunas instituciones (p. 23).

A partir de lo anterior, si bien no existe nada inherentemente incorrecto o perjudicial en liberar información sobre las personas, las preferencias de estas varían de manera drástica en cuanto a lo que debe o no debe ser privado (European Data Portal, 2016), por lo que es difícil poder discriminar entre las opciones posibles. Por esta razón, cualquier listado de reglas posibles para proteger la privacidad puede ser considerado muy simplista al momento de alinearse con las preferencias reales de las personas (European Data Portal, 2016). Sin embargo, es necesario contemplar un conjunto de factores base en cualquier investigación, incluyendo peligros, sensibilidad, normas, confidencialidad e interés público. Los peligros potenciales son difíciles de cuantificar, pero el peligro más esperable es que afecte la reputación de una persona o institución. En cuanto a la sensibilidad, ciertos tipos de información son más sensibles que otras, por ejemplo, la salud, la orientación política, el estado financiero, etc. Las normas refieren a la normativa presente en cualquier sociedad de la que la investigación debe hacerse cargo. Asimismo, cuando se ha estipulado la confidencialidad no es prudente hacer pública la información, aunque en este acto no se esté rompiendo el principio de privacidad, ya que no necesariamente se está liberando información acerca de los individuos. Y, por último, el interés público puede ser de gran utilidad, la que puede superar los intereses privados que mantienen la información en secreto; por esto, la relevancia de estar contextualizado a un conjunto de normas (European Data Portal, 2016).

3. Desarrollo histórico de la implementación de datos abiertos

Contexto internacional

En el año 2003, se desarrolló la agenda digital europea con la finalidad de poner en marcha una estrategia de datos abiertos en los países miembros. En esa ocasión Neelie Kroes, vicepresidenta de la Comisión Europea, declaró lo siguiente: “Enviamos hoy una señal clara a las administraciones. Sus datos valen más si se difunden. Háganlo entonces ahora: aprovechen este marco para sumarse a otros dirigentes inteligentes que ya están sacando ventaja de los datos abiertos. Los contribuyentes ya han pagado esta información y lo menos que podemos hacer es devolvérsela a quienes deseen usarla de nuevas maneras que ayuden a las personas y generen crecimiento y puestos de trabajo” (European Commission, 2003). Como objetivo de la agenda, el énfasis recae en la reutilización de los datos gubernamentales, del patrimonio y de los científicos (Alcocer, Blasco-Gil & Peset, 2013).

En esta ocasión, se concibe el concepto de datos abiertos como a aquel referido a un subconjunto de datos que se ponen libremente a disposición de cualquiera para su reutilización tanto para fines comerciales como no comerciales (Unión Europea, 2014, p. 5).

De la misma manera, el año 2009, el entonces presidente de Estados Unidos, Barack Obama, desarrolló un proyecto de apertura a la información para la administración de su gobierno y generó una política de acceso señalando que “un solo bit de información puede generar un torrente de creatividad” (Solana, 2011). Con esta concepción, su administración se comprometió a crear un nivel de apertura a la información desconocido hasta el momento, que buscaba asegurar la confianza pública y establecer un sistema de transparencia, participación pública y colaboración, teniendo como resultado final fortalecer la democracia y promover la eficiencia y efectividad en el gobierno (Obama, 2009). Se enfatizó, entonces, en tres características que debe tener un gobierno abierto: a) debe ser transparente, b) debe ser participativo, y c) debe ser colaborativo (Obama, 2009).

Acorde al proceso iniciado en 2003, que fue continuamente desarrollándose en los años siguientes, en 2013 se redactó una carta internacional de datos abiertos que contenía una visión conjunta y de alto nivel para el desarrollo de iniciativas de los países miembros del G8. Esta iniciativa fue el primer esfuerzo multilateral por introducir principios comunes internacionales para los datos abiertos (Gobierno abierto, 2015). Su objetivo fue dar mayor coherencia y establecer una mejor colaboración para el acuerdo en los principios, estándares y buenas prácticas que deben regir los datos abiertos en el mundo, con la aspiración de continuar desarrollando recursos, guías y herramientas para promover la adopción de los principios con el mayor número posible de agentes (Open Data Charter, 2015).

Es en el documento señalado en el que se expresó, por primera vez, de manera colaborativa e internacional, la intención de consensuar los “datos digitales, entendiéndolos como aquellos datos que son puestos a disposición con las características técnicas y jurídicas necesarias para que puedan ser usados, reutilizados y redistribuidos libremente por cualquier persona, en cualquier momento y en cualquier lugar” (Open Data Charter, 2015, p.5). A este respecto, se considera que los datos abiertos están en el centro de un cambio global que refiere a una transformación tecnológica de medios digitales impulsada por los datos y la información, y de la cual los datos abiertos son cruciales para cumplir los objetivos de una sociedad más próspera, equitativa y justa. Así, estos datos entregan una oportunidad única que el mundo global debe aprovechar, porque presentan oportunidades para encontrar soluciones con políticas innovadoras y basadas en la evidencia, y fomentan beneficios económicos y desarrollo social para quienes integran la sociedad (Open Data Charter, 2015).

Asimismo, los países participantes de la carta internacional de datos abiertos reconocen la potencialidad y la magnitud de los datos disponibles que son de interés de la ciudadanía pero que, sin embargo, no son aprovechados en su totalidad. Se entiende así que abrir y hacer disponible la información promueve la construcción de sociedades más interconectadas, e incentiva el desarrollo de la innovación, de la justicia, de la transparencia y de la prosperidad en la medida que impulsa la participación cívica en la toma de decisiones públicas.

En consecuencia, el acuerdo internacional refiere a seis principios que forman la base del acceso a los datos para su publicación y uso, y que actualmente siguen vigentes. Estos son (Open Data Charter, 2015):

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

Asimismo, el Consejo Europeo, llevado a cabo también en el año 2013, hizo explícita la importancia de la economía digital, de la innovación y de los servicios como motores del crecimiento y del empleo. De esta manera, los países miembros solicitaron la generación de un marco en el que se delimiten las características de un mercado de macrodatos (big data) y la computación en nube (cloud computing) (Comisión Europea, 2014).

En el marco de encontrarnos en un contexto mundial de datos voluminosos y complejos (...) las actividades humanas, los procesos industriales y la investigación conducen a la recolección y el procesamiento de datos en una escala sin precedentes, impulsando nuevos productos y servicios, así como nuevos procesos de negocio y metodologías científicas (Unión Europea, p. 2, 2014).

Con esto, los conjuntos de datos son tan voluminosos y complejos que la sociedad debe hacerse cargo de procesarlos y hacerlos disponibles. En un contexto global de digitalización acelerada de los servicios públicos, se impulsa la “necesidad de modernizar, reducir costes y ofrecer servicios innovadores, y abre nuevas oportunidades para optimizar el almacenamiento, la transferencia, el procesamiento y el análisis de datos” (Unión Europea, 2014, p. 2).

A este respecto, los países miembros de la Unión Europea y Estados Unidos comenzaron un plan de acción para el desarrollo y publicación de datos abiertos, lo que tuvo como consecuencia la generación de estrategias para el desarrollo tecnológico e investigativo por parte del resto de los países del mundo que se encuentran desplegando y avanzando en sus sistemas de datos abiertos.

Reflejo de este proceso es el informe del Barómetro de Datos Abiertos ejecutado por la Web Foundation. En su primera edición, el año 2013, el Barómetro encontró que ha existido una difusión importante en las políticas de información de las que los gobiernos se han hecho cargo de maneras diferentes. Mientras que la mayoría de los países se encuentra en estadios bajos de desarrollo respecto al acceso a la información, todos están explorando diferentes formas de hacer la información disponible y estimular la innovación y la transparencia de los datos (Web Foundation, 2013). A pesar de esto, la cuarta edición y más actual del Barómetro, elaborada en 2017, que abarcó 115 países y jurisdicciones, concluyó que los gobiernos están avanzando a un menor ritmo e incluso algunos han interrumpido su compromiso con los datos abiertos (Web

Foundation, 2017). De esta manera, o la información que estuvo disponible ya no está vigente o no es suficiente la cantidad y calidad de los conjuntos de datos publicados. Sin embargo, existen países que han adoptado formalmente la Carta de Datos Abiertos y han conseguido buenos avances respecto del cumplimiento de los principios del acceso a la información, entre ellos Chile, que se encuentra en el número 23 de la clasificación del Barómetro. Respecto al contexto latinoamericano, el Barómetro identificó a México (11) como el líder de la región seguido por Uruguay (17) y Brasil (18).

Contexto nacional

En este contexto, en 2008 se promulgó en Chile la Ley Nº 20.285 sobre acceso a la información pública, que se conoce como Ley de Transparencia. Entre las menciones a las que hace referencia esta ley, se expresa que la función pública se ejerce con transparencia, de modo que permite y promueve el conocimiento de los procedimientos, contenidos y decisiones que se adopten en ejercicio de ella. De acuerdo con esto, se entiende que el principio de transparencia de la función pública consiste en respetar y cautelar la publicidad de los actos, resoluciones, procedimientos y documentos de la administración, así como la de sus fundamentos, y en facilitar el acceso de cualquier persona a esa información mediante los medios y procedimientos que al efecto establezca la ley. Asimismo, la ley presenta la obligatoriedad del acceso público a la información perteneciente a cualquier organismo público.

Es pública la información que obre en poder de los órganos de la administración, cualquiera sea su formato, soporte, fecha de creación, origen, clasificación o procesamiento, a menos que esté sujeta a excepciones (Artículo 5º, Ley Nº 20.285).

Si bien esta ley abarca la información contenida en cualquier organismo público y no hace distinción respecto al carácter de la información (administrativa, planificación, presupuesto, investigación, etc.), entrega el punto de inicio para, desde el ámbito gubernamental, estar regidos como instituciones públicas a la transparencia de la información; se incluye en este dictamen a todo lo referido al ámbito de la investigación. A pesar de esto, si bien se propende a la apertura de la información desde el sector público, también se restringe su uso y difusión por medio de la Ley Nº 19.628 Sobre protección de la vida privada en que se define que:

En toda recolección de datos personales que se realice a través de encuestas, estudios de mercado o sondeos de opinión pública u otros instrumentos semejantes, sin perjuicio de los demás derechos y obligaciones que esta ley regula, se deberá informar a las personas del carácter obligatorio o facultativo de las respuestas y el propósito para el cual se está solicitando la información. La comunicación de sus resultados debe omitir las señas que puedan permitir la identificación de las personas consultadas. El titular puede oponerse a la utilización de sus datos personales con fines de publicidad, investigación de mercado o encuestas de opinión (Artículo 3º, Ley Nº 19.628).

En este marco, el gobierno definió, en el año 2012, como un eje fundamental en su gestión, el impulso de una agenda de modernización de las instituciones públicas con la finalidad de ser un gobierno al servicio de su ciudadanía. Así, el desafío fue fortalecer la democracia haciendo que la transparencia, la participación y la colaboración fueran accesibles por la ciudadanía (Piñera, 2012).

Surge entonces el proyecto Datos.gob.cl como iniciativa del Gobierno de Chile, enfocado en promover el acceso expedito, abierto y sin restricciones de uso a conjuntos de datos gubernamentales que permitan su reutilización por parte de terceros; se entiende que pueden agregar valor a la información disponible y potenciar la divulgación e incorporación de información que permita el desarrollo de políticas, programas, etc. de relevancia para el país (Minsegespres, 2013). De esta manera, el portal permite centralizar en una plataforma unificada las iniciativas de transparencia, participación y datos públicos, generando estándares en las instituciones públicas, facilitando su cumplimiento y simplificando la comprensión y el uso por parte de la ciudadanía (Piñera, 2012).

En este modelo de datos abiertos, el Minsegespres (2012) expone que las diferentes instituciones gubernamentales deben poner a disposición de toda la población los datos que han sido generados por cada institución. A este respecto, es necesario, al abordar la preparación y exposición de datos públicos en formato electrónico en Chile, normalizar los procesos para que cada institución haga uso de metodologías, procedimientos y prácticas comunes que garanticen la organización lógica, física y de formatos de los datos publicados. Por consiguiente, la conceptualización de datos abiertos no es solamente una decisión técnica sino también política, que busca igualar el acceso a la información, garantizar los derechos de la ciudadanía y reducir las brechas de desigualdad, y generar diálogo, participación, acceso a la información y espacio para el aporte de la ciudadanía en la toma de decisiones de asuntos públicos (Minsegespres, 2016).

El caso del Ministerio de Educación

Para impulsar el desarrollo de los datos abiertos, el Ministerio de Educación promueve la función del Departamento de Estudios y Desarrollo (Centro de Estudios) cuya función es producir, promover y difundir información y conocimiento respecto del sistema educativo. Con tal objetivo, dicho departamento cumple diferentes funciones y las principales son: a) la generación de estadísticas, información, estudios y conocimiento como aporte al diseño, implementación y evaluación de las políticas, programas y proyectos ministeriales en materia educativa; b) el apoyo y articulación de diversos actores e instituciones nacionales e internacionales en la realización de informes, reportes y estudios en el ámbito educativo; c) la promoción de Investigación y Desarrollo [I+D] con impacto en políticas educativas; y d) la difusión de estadísticas, informaciones, evaluaciones y estudios relevantes a la comunidad educativa (Centro de estudios, 2018).

De manera concreta, el Centro de Estudios, en su función de ser la entidad dentro del Ministerio de Educación que desarrolla y apoya la ejecución de diferentes estudios y documentos, ha implementado dos plataformas de datos desde las cuales pone a disposición pública un compendio de las principales informaciones y documentos que se presentan con acceso libre por

medio de una página web. Con ellas, se busca promover la transparencia, participación y colaboración ciudadana. En primer lugar, se materializó una plataforma web, de fácil acceso y navegación, comprensible y utilizable por cualquier persona interesada, que reúne estadísticas de la educación. En segundo lugar, está disponible el Centro de Documentación (CEDOC) que es un repositorio del conjunto de estudios, evaluaciones, reportes y publicaciones ejecutadas en y para el Ministerio de Educación. Ambas iniciativas buscan democratizar el acceso a datos publicando gran cantidad de información del sistema educativo.

A pesar de la existencia de las plataformas mencionadas, si bien estas son reflejo del esfuerzo por hacer disponible información de fácil acceso y potencial reutilización, existen características que las hacen aun incompletas. Al respecto se observa, por ejemplo, que la plataforma de publicaciones CEDOC responde a un repositorio de información final y no contiene almacenamiento de material complementario. Asimismo, si bien es de acceso libre, tanto el formato de los documentos como su disposición dentro de la página web están diseñadas para un uso de una comunidad específica que conoce su empleo (en este caso podrían ser investigadores, investigadoras, funcionarias y funcionarios del mismo ministerio).

Específicamente, se puede dar cuenta de las características de cada plataforma respecto de la conceptualización de datos abiertos. Para este caso, si bien la plataforma CEDOC es accesible en cuanto posee documentos de libre acceso, que se mantienen en el tiempo y que son potencialmente reutilizables en cuanto pueden ser utilizados para referencias bibliográficas (pero no para el empleo de los datos en concreto), estos no son localizables ni interoperables. Manifiesto de esto es la encuesta de percepción de la plataforma realizada por el Centro de Estudios (2018) en que el contenido de la página web fue calificado por la mayoría como insuficiente (48%) y poco claro (51%), a la vez que se evaluó la experiencia de navegación como desordenada (58%) y con un diseño poco atractivo (42%).

En cuanto a la plataforma de información estadística Datos Abiertos, si bien la información presentada cumple con las cuatro características que deben lograrse, no se encuentran completamente aplicadas. Ejemplo de esto es la información faltante y que, por lo tanto, no está completa; es necesario facilitar aún más su acceso y posible utilización, y agregar más información no solo estadística para complementar la información presentada.

En consecuencia, se entiende que el acceso a la información se encuentra en un proceso de transformación, en que si bien existe una amplia disponibilidad de datos de acceso libre, aún se encuentra en un proceso de desarrollo hacia datos abiertos.

4. Problema de investigación

El aumento de datos disponibles y la gran velocidad con que estos se van generando en la actualidad evidencian un panorama en el que existen muchas fuentes y tipos de datos disponibles, un gran volumen de ellos, mucha volatilidad y poco uso de la información, sin existir finalmente estándares de calidad asociados (Ferrer-Sapena, 2017).

En este contexto es que el actuar de un centro de investigación, al estar relacionado con la producción de datos, debe también estar enmarcado en el uso y acceso a ellos. En este sentido, los datos solo son útiles en la medida que se utilizan para generar conocimiento para la acción, por lo que cuando se habla de datos masivos también se está hablando de gestión de la información y generación de conocimiento para la acción (Ferrer-Sapena, 2017). A este respecto, para el desarrollo de políticas públicas educativas, se vuelve necesario que el Centro de Estudios del Ministerio de Educación abra los datos para que una mayor parte de la ciudadanía pueda aplicar esa información para la generación de conocimiento.

Para esto se requiere generar estándares de calidad no solo en el desarrollo de la investigación, proceso que actualmente se garantiza mediante el acompañamiento y evaluación de todos los productos e investigaciones realizadas por equipos externos al Ministerio, sino también se necesita establecer estándares que permitan que dicha información se encuentre disponible y sea localizable, accesible y reutilizable.

Así, el problema de investigación del presente es identificar cuáles son los fundamentos, las condiciones y los estándares necesarios para el registro, almacenamiento y publicación de datos e información de calidad.

De esta manera, se pretende elaborar estándares para la generación de la información y así homogeneizar la información obtenida para potenciar su utilización y reutilización. Finalmente, se busca promocionar el uso de la información por parte de personas internas y externas al Ministerio, e instar también la colaboración entre actores de la comunidad educativa, evitando duplicación de procesos de investigación y permitiendo, a su vez, validar los materiales de investigación realizados (Ferrer-Sapena, 2017).

5. Metodología

Dado que el presente documento aborda las condiciones y estándares necesarios para el registro, almacenamiento y publicación de los datos abiertos de gobierno, específicamente en instituciones públicas, se aplicó una metodología de observación que permitió describir, de manera exploratoria, cómo se realiza la publicación de información con fines investigativos en organismos públicos equivalentes al Ministerio de Educación. El objetivo no fue evaluar la calidad ni cantidad de los datos disponibles, así como tampoco calificar el tipo de soporte tecnológico en que están

disponibles, sino que identificar iniciativas de publicación de datos que cumplan con las características básicas de los datos abiertos.

Los criterios básicos observados fueron tomados de la conceptualización generada por la Open Knowledge Foundation (2012), la cual especifica que los datos deben cumplir con ser accesibles (de acceso libre sin restricción para ser empleados por cualquier persona), potencialmente utilizados (estar disponibles por medio de internet y en un formato que permita su modificación) y, reutilizados y redistribuidos (en condiciones óptimas para combinarse con otros conjuntos de datos), permitiendo la creación de obras derivadas de los mismos datos.

Al respecto, se generó como instrumento de recolección, una pauta estructurada que contempló 28 preguntas cerradas y 1 pregunta abierta construidas en torno a tres dimensiones asociadas a los tres criterios antes mencionados:

- Disponibilidad, que refiere a que los datos sean potencialmente utilizados en la medida en que estén disponibles en las plataformas online;
- Acceso, que refiere al nivel de accesibilidad que poseen los datos en tanto se encuentran libres de restricciones y el formato permita ver su contenido;
- Reutilización, dimensión referida a la posibilidad que tiene la información de ser potencialmente reutilizada en futuras investigaciones. De manera concreta, esto se observa en la disponibilidad de datos primarios como bases de datos, transcripciones, audios y otro tipo de información.

La aplicación de este instrumento contempla el carácter público desde el cual se está ejecutando la observación. Al respecto, y entendiendo que la información a observar refiere a los datos abiertos de gobierno, específicamente en lo referente a datos de investigación, y que la información perteneciente a este sector es muy amplia y contempla diferentes áreas de conocimiento, es necesario identificar qué se comprende, para objetivos de este documento, como datos abiertos en este contexto.

Al respecto, Torres, Robinson & Cabezas (2012) consideraron como dato a todo “material que ha sido registrado durante una investigación, reconocido por la comunidad científica y que sirve para certificar los resultados de la investigación que se realiza” (p.7), la cual debe provenir de una fuente única siendo estos difícil de obtener nuevamente. Estos datos pueden incluir cuadernos de laboratorio y de campo, datos de investigación primaria, transcripciones, bases de datos, cintas de audio, videos, desarrollo de modelos, fotografías, entre otros.

Muestra

Dado el carácter del Centro de Estudios, en tanto organismo que forma parte de la comunidad de investigación, por un lado, y organismo público que es parte de la Administración del Estado, por otro, el universo de revisión generado es la totalidad de los organismos públicos de la administración central de Chile. Esto corresponde a 23 ministerios y 156 servicios públicos. Sin

embargo, dado el tiempo y la capacidad de revisión se optó por trabajar con una muestra intencionada que incluyó solo a los 23 ministerios de la República.

De esta manera, se analizaron las plataformas web de los ministerios² especificados, teniendo como foco principal la página o sección que contiene la información estadística y/o los estudios y/o investigaciones. En esta revisión no fue incluida la revisión de páginas de estudios y/o estadísticas de organismos asociados o dependientes de los ministerios, sin embargo, en aquellos casos donde los estudios y/o estadísticas se encontraban radicados en sus respectivas subsecretarías, estos sí fueron incluidos en la revisión³.

La observación consideró el análisis todos aquellos espacios donde fuera factible encontrar conjuntos de información pública susceptible de ser utilizados, esto incluye desde repositorios de información (correspondiente a la vía verde de datos abiertos) hasta portales de datos abiertos. Por otra parte, la revisión consideró también como dato la información procedente de registros administrativos⁴ de cada ministerio. Según Trias (2013), los datos administrativos usualmente son recolectados por tres razones: 1) para el seguimiento de programas, 2) para la focalización de programas sociales y 3) para cumplir con regulaciones y ejecutar auditorías. Estos datos tienen un gran potencial investigativo, ya que, por lo general, son recogidos de manera frecuente y poseen una amplia cobertura.

Por último, respecto a las limitaciones de esta investigación se identificó que al usar las páginas web como fuente de información no es posible saber con certeza las características completas de los datos y, por tanto, su capacidad de utilización y reutilización. Esto en referencia a que la publicación de datos puede responder a una selección de la información que se cree relevante o publicable y, por ende, no es necesariamente la información completa. De esta manera, es de esperar que existan más estudios, investigaciones y estadísticas de los que se encuentran disponibles vía online en todos los ministerios, sin embargo, la observación solo puede realizarse con lo que se encuentra disponible y que es el material que cada institución considera como relevante de difundir.

6. Resultados

En esta sección se presentan los resultados de la manera en que las instituciones gubernamentales abordan la apertura de la información con potencial uso para fines investigativos. Tal como se mencionó, el objetivo no es evaluar a las instituciones sino conocer cómo se desarrolla la publicación de información con posibles fines investigativos para mejorar el trabajo que efectúa el

² Ver Anexo 1: Listado ministerios y sitios web analizados.

³ Si más de una subsecretaría poseía página, si bien se aplicó la pauta a cada página, los resultados son agregados por Ministerio.

⁴ Un registro administrativo se define como una serie de datos sobre una acción sujeta a regulación o control obtenida por una entidad pública o privada como parte de su función, y que es resultado de necesidades fiscales, tributarias u otras, creada con la finalidad de visibilizar la administración de los programas de gobierno o para fiscalizar el cumplimiento de obligaciones legales de la sociedad (DANE, 2010).

Ministerio de Educación, por lo tanto, la información se presenta de manera agregada, sin individualizar a los ministerios analizados, y buscando relevar las experiencias y no a las instituciones.

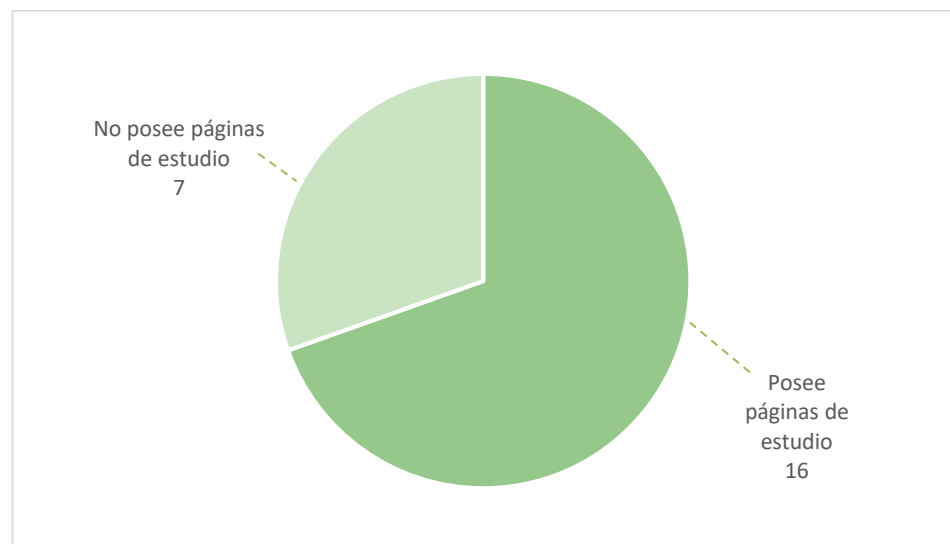
La información se presenta en torno a las tres dimensiones identificadas: disponibilidad, acceso y reutilización.

Disponibilidad

En relación con la disponibilidad, esto es, la existencia y el tipo de datos disponibles en los sitios web de los ministerios de la República de Chile, se observó que 16 de los 23 poseían al menos un espacio en sus páginas oficiales para difundir estadísticas, estudios o investigaciones.

Hubo siete ministerios que no fueron incluidos en los análisis debido a diversos motivos. En el caso de dos ministerios (y sus respectivas subsecretarías), al momento de la revisión no disponían de páginas propias para la difusión de investigaciones y estadísticas, y esto era ejecutado por organismos dependientes que desarrollan esta labor de manera autónoma. Hay otros tres ministerios que si bien poseían un repositorio de documentos, no necesariamente correspondían a estudios ni investigaciones, por lo tanto, el tipo de datos no permitía incluirlos en el análisis. Finalmente, uno de los casos no tenía en funcionamiento la sección de estudios y/o estadísticas, y, en otro caso, la información que se entregaba correspondía solo a mapas e información georreferencial.

- No disponen de páginas (2)
- Tienen repositorios de documentos que no son informes de investigación (3)
- Página web no está activa (1)
- Exclusivamente información georreferencial (1)



En los 16 ministerios donde fue posible identificar un sitio web asociado a estudios y/o estadísticas, se revisó el tipo de información que se pone a disposición de la comunidad. Se distinguió entre dos categorías: informes de investigación y datos con potencial para la investigación.

En todos los casos observados se encontraron informes de estudios e investigaciones, que evidenció una vía verde de los datos abiertos en estos ministerios, probablemente impulsado o potenciado por la implementación de la Ley de Transparencia. Cabe recordar que esta revisión no permitió identificar el porcentaje de estudios, investigaciones y datos (cualitativos y cuantitativos) que existen en cada una de las reparticiones, sino solo las disponibles en la plataforma ministerial.

En cuanto a los datos con potencial investigativo se buscaron dos tipos de datos: cuantitativos y cualitativos. En este caso, lo que se revisó fue la existencia de este tipo de datos y su origen, es decir, si estaban producidos a partir de investigaciones o si se producían a partir de procesos de orden administrativo/normativos. En relación con la información cuantitativa, 10 de los 16 organismos disponían de bases de datos que permitían la generación de estadísticas del sector. En tres casos estas bases procedían de registros administrativos y encuestas; en cuatro casos las estadísticas fueron generadas a partir de datos administrativos; y en tres, solo a partir de encuestas. En cuanto a los datos cualitativos, no fue posible encontrar ningún organismo que dispusiera de manera online de datos cualitativos, ya sea transcripciones de entrevistas individuales o grupales, audios, videos u otro material de estas características.

Aunque no estaba contemplado en el instrumento de recolección original, la disponibilización de información georreferenciada surgió como un tema emergente que ha sido incorporado en algunos organismos públicos.

Acceso

Esta sección hace referencia al nivel de accesibilidad que poseen los datos. En primer lugar, se indagó en la facilidad para acceder a la información, según la cantidad de sitios dentro de un ministerio que disponen de datos (lo que da una referencia de cuán dispersa se encuentra la información) y en la existencia de buscadores que permitieran una exploración más especializada. Además, se revisaron los formatos disponibles, esto es, si los datos se encontraban disponibles en más de un formato lo que minimiza las restricciones al usuario en términos de licencias de software y, por último, se hizo referencia a las restricciones en términos de acceso al dato, es decir, si se establecían mecanismos de registro o de autenticación para acceder a determinada información.

El primer elemento para conocer cómo es el acceso a la información de los organismos públicos se relaciona con la facilidad para acceder a los sitios web que entregan datos. En este sentido, se pudo identificar que la información de un mismo ministerio no necesariamente se encontraba en un único sitio web. En ocho de estos organismos, es decir, la mitad de los casos analizados no existía una única página web, y la información se encontraba en sitios independientes a pesar de

que pertenecían al mismo servicio. En estos casos los datos estaban divididos por subsecretarías, por divisiones o departamentos; en otros casos, a raíz de la envergadura del proyecto (por ejemplo, encuestas nacionales), se generaban páginas independientes que funcionaban de manera paralela a las páginas de estudios o estadísticas.

Por otra parte, la presencia de buscadores catalogados en función de palabras claves u otros descriptores de búsqueda no es un recurso muy empleado en los organismos públicos. Únicamente ocho instituciones tenían buscadores especializados, en el resto de los casos los buscadores eran simples (solo en base a ingreso de palabras) o no tenían ningún tipo de buscador que facilitara encontrar la información.

Según la literatura relacionada con datos abiertos, el acceso también refiere a las opciones de formato en que se disponibilizan los datos en tanto se amplía el potencial de usuarios que pueden acceder a ellos. En el caso de los reportes de investigaciones o estudios, la mayoría se encuentra en un único formato, sin entregar alternativas. En los reportes de investigaciones, el formato preferente es PDF y lo sigue el PPT.

En cuanto a las bases de datos, en siete de las 16 instituciones se pudo detectar al menos una base de datos que se encontraba disponible en más de un tipo de formato, siendo el formato preferido XLS o XLSX, seguido por CVS y un tanto más atrás, formatos como SAV o STATA. En el total de los casos revisados solo se identificaron dos ministerios que disponen sus datos con formatos API (formato para lectura por máquinas) que permiten el trabajo para desarrolladores.

En relación con las restricciones para acceder a la información mediante requisitos de registro, hay tres ministerios en que se solicita registro o información adicional para acceder a ciertos datos. No obstante, dado que no es posible conocer cuánto del total de información de cada organismo está disponible de manera online, este es un dato que debe ser interpretado con precaución debido a que otros datos podrían estar siendo solicitados mediante mecanismos de transparencia u otros. Además, estas solicitudes de información podrían estar vinculadas al tratamiento de datos sensibles.

Reutilización

Esta dimensión se encuentra asociada a la posibilidad que tiene la información de ser reutilizada en futuras investigaciones, para lo cual se requiere contar con datos primarios como bases de datos, transcripciones, audios y otro tipo de información. Junto con esto, para la reutilización también es necesario que estos datos cumplan con ciertas condiciones que permitan su posterior uso. Este trabajo identificó si los organismos disponían de datos con formatos estandarizados (es decir, con un único formato para la presentación de la información) y si contaban con variables o información que permitieran la interoperabilidad, es decir, que posibiliten la interrelación entre conjuntos de datos, por ejemplo, mediante códigos de identificación de casos. Cabe recordar que el análisis se hizo por ministerio y no por cada conjunto de datos.

Tal como se señaló en la sección de disponibilidad, los datos susceptibles de ser empleados en investigación que están disponibles de modo abierto son exclusivamente de tipo cuantitativo y georreferencial, y no fue posible encontrar datos cualitativos.

En cuanto a los datos cuantitativos, se pudo identificar cinco ministerios que tienen la totalidad de sus bases de datos publicadas online con un único formato, es decir, hay un grado mínimo de normalización en la entrega de información. Respecto a la interoperabilidad, se midió mediante la existencia de alguna variable en una o más bases de datos del organismo que permitiera interrelacionar ese conjunto de datos con otros. En este caso, fueron identificados tres ministerios que incluían este tipo de variables en al menos una de sus bases de datos públicas.

Respecto de la reutilización de datos procedentes de investigaciones se identificó que solo seis de los 16 ministerios que poseen sitios web asociados a estadísticas y/o estudios, acompañaban los informes de investigación con información adicional. De los seis ministerios que entregan información adicional al informe final, solo uno lo incluía como práctica para la totalidad de las investigaciones publicadas. En el resto la publicación de los subproductos es discrecional dependiendo de la investigación, es decir, no en todas las investigaciones están disponibles.

De los seis ministerios que sí publican información asociada a los informes de investigación, en todos corresponden a bases de datos, en cuatro ministerios se publican los instrumentos, en tres se entrega información de índole administrativa (términos de referencia, contratos o similares) y en dos ministerios se publican los informes finales. En ningún caso se publican transcripciones completas, audios o videos asociados a investigaciones, es decir, datos primarios de investigaciones de corte cualitativo⁵. La publicación de datos investigativos es más usual en aquellos estudios de alcance nacional, principalmente encuestas institucionales y regulares (como CASEN o Encuesta Longitudinal de Empresas), y es menos frecuente en investigaciones de menor alcance.

Al interior de cada ministerio existen diferencias según el informe de investigación en el tipo de subproductos que se ponen a libre disposición. En algunas investigaciones dentro un mismo organismo se publican los términos de referencia y contratos, en otros casos, se publican los informes parciales o presentaciones asociadas al estudio y, en otros, la totalidad de subproductos de la investigación, esto es, instrumentos, manuales y bases de datos. Pero no fue posible encontrar ningún ministerio que tuviera un estándar que regule los productos publicables.

Dada esta situación, si bien hay prácticas orientadas a publicar contenidos asociados a la realización de investigación, no todo el material publicado tiene potencial de ser concebido como dato abierto, específicamente, el material asociado a actos administrativos (contratos, términos de referencia, entre otros) responde más bien a la perspectiva de gobierno abierto impulsado por las políticas de transparencia y no a la reutilización de los datos, lo que reduce aún más el número de experiencias que apuntan en esta dirección.

⁵ Cabe señalar que el diseño no contemplaba la revisión del contenido de los informes, por lo tanto, si algún subproducto fue incorporado como anexo dentro de los informes esto no pudo ser detectado.

7. Conclusiones

El gran volumen de datos e información que se está generando alrededor del mundo ha impulsado una gran cantidad de conceptualizaciones y recomendaciones de cómo difundirlos y emplearlos. A pesar de que se identifican matices en cómo se operacionaliza la definición de datos abiertos y de cómo debe ser evaluado y medido, existe un consenso en la necesidad de generar las condiciones para el reuso o reutilización de la información. En especial, en el ámbito público se identifica el gran potencial de la apertura de datos como un elemento democratizador que permite la generación de evidencia para la política pública.

La promulgación de la Ley de Transparencia el año 2008, y el desarrollo de la iniciativa de gobierno en 2012 que promueve el acceso expedito, abierto y sin restricciones de la información son reflejo de esfuerzos del Estado de Chile por unirse a la revolución de la información que ocurre durante los últimos años a nivel mundial.

Al respecto, y luego de unos años de la primera iniciativa nacional, se puede observar que al año 2018 son dispares los avances que se han llevado a cabo en los diferentes ministerios, coexistiendo experiencias que se encuentran en diferentes niveles de progreso y que presentan desarrollos en áreas diversas de la información, identificándose entonces distintos estadios entre ellos. Por una parte, existen aquellos cuya entrega de información está limitada al cumplimiento de la normativa vigente ⁶, que les exige poner a disposición información administrativa de la institución y recibir solicitudes de información por parte de personas naturales. Por otra, se encuentran aquellos que además de poner a disposición la información exigida por transparencia están desarrollando un proceso activo de apertura de información a la comunidad, en la que se publica información sobre diferentes aspectos desarrollados por el organismo y que buscan contribuir con más información pública y con potencial de utilizarlo para investigación; este panorama refleja de manera más fiel un proceso de datos abiertos de gobierno.

Con todo, a partir de la revisión de la situación de los ministerios de la República de Chile se puede concluir que:

- a) El concepto de dato aún tiende a ser considerado de manera acotada y muy vinculada al dato cuantitativo, pudiéndose también identificar que -de manera emergente- están teniendo cada vez más peso y notoriedad los datos georreferenciados. No se observa la presencia de datos de origen cualitativos (transcripciones, audios, videos) debido, probablemente, a las dificultades que se deben mantener en la protección del derecho de las personas (en especial en videos).
- b) Asimismo, los datos disponibles -factibles de ser ocupados en investigación- proceden en su mayoría de registros administrativos y encuestas nacionales regulares e institucionalizadas. No obstante, en investigaciones de menor magnitud o no regulares

⁶ Ley de Transparencia.

no existe como práctica institucionalizada disponer en forma libre y sin restricciones de datos primarios asociados a las investigaciones ejecutadas con recursos públicos.

- c) Del mismo modo, tampoco es una práctica institucionalizada hacer disponibles otros materiales asociados a las investigaciones como los instrumentos mediante los cuales se recoge la información, los materiales de registro del terreno de la investigación, y los documentos intermedios de avance, etc., que complementan y aportan al quehacer investigativo.

A este respecto, el Ministerio de Educación, por medio de la información que pone a disposición el Centro de Estudios, se encuentra comparativamente en estados avanzados respecto de los demás ministerios, dado que, por un lado, la información disponible no solo corresponde a la vía verde de los datos abiertos (concretado en el repositorio CEDOC) sino que ha progresado en el desarrollo de un sitio web que posibilita la reutilización de la información y destaca tanto en la cantidad de bases de datos disponibles como en la estandarización de estas.

Es posible aseverar entonces que el Ministerio de Educación se encuentra alineado con el contexto internacional y se ha hecho cargo de la responsabilidad instaurada por la Ley de Transparencia y la iniciativa de gobierno en 2012 respecto del acceso público a la información. Específicamente, los datos publicados por el Centro de Estudios cumplen con criterios especificados en el marco teórico y en la metodología en la medida que son disponibles (están de manera permanente en las plataformas y son posibles de descargar), accesibles (en la medida en que se encuentran disponibles distintos formatos) potencialmente interoperables (en la medida que están en un formato modificable y que dispone de variables que permiten potencialmente complementarlo con fuentes de información internas y externas al ministerio) y reutilizables (se presentan en formato adecuado y tienen la información suficiente para poder ser empleadas en diferentes investigaciones).

No obstante, tras la revisión de la literatura y de los resultados de este estudio hay espacios de mejora que son factibles de abordar, no solo en el Ministerio de Educación, sino en las distintas instituciones del Sistema de Aseguramiento de la Calidad.

- a) Aunque las plataformas tienen disponibles informes finales de investigación y estadísticas, usualmente no presentan materiales complementarios al proceso de investigación (datos primarios, instrumentos, materiales de registro del terreno de la investigación, entre otros) que complementen y aporten al quehacer investigativo. Dado esto, un espacio de mejora lo constituye disponer no solo de los informes finales de investigación, sino que de los datos primarios e información complementaria que permita la reutilización de estos datos.
- b) Avanzar hacia la consideración como datos abiertos de distintos tipos de datos no solo cuantitativos sino también cualitativos, dado el potencial comprensivo que posee esta información y su potencialidad para comprender los fenómenos sociales.
- c) Lo anterior requiere mínimamente definir estándares para el registro y el almacenamiento de datos de investigación (de distinta índole) producidos con recursos

públicos. En la medida que se avance hacia la normalización de la forma en que se registran y almacenan estos datos, se aumenta la posibilidad de convertir esta información en datos abiertos que cumplan con las condiciones establecidas en la literatura.

- d) A su vez, esto requiere repensar las plataformas disponibles para la entrega de datos; en el caso del Ministerio de Educación, la información accesible se encuentra disponible en dos plataformas que dividen el contenido entre estadísticas, por un lado, e investigaciones y documentos desarrollados por el Centro de estudios, por otro. A partir de la relevancia que se le da a la facilidad de acceso a la información, es considerar la realización de una plataforma que contenga toda la información referente a las dos plataformas existentes, incorporando los materiales complementarios al proceso de investigación mencionados en el punto anterior, o bien, una interoperabilidad entre ambas plataformas.

La disponibilidad de mayor cantidad de datos primarios cuantitativos y cualitativos es un potencial para la generación de más conocimiento en educación, pero este no es el único beneficio de implementar un enfoque de datos abiertos en el Sistema de Aseguramiento de la Calidad. También podría contribuir el evitar la sobrecarga de demandas a las comunidades de los establecimientos educativos, pues en la medida que se publica información confiable y con posibilidad de reutilización, se podría soslayar la duplicación de esfuerzos en la recogida de bases de datos, focalizando el esfuerzo en el análisis e interpretación de estos.

De acuerdo con lo mencionado y en relación con las vías de acceso en las que se disponibilizan los datos, se puede evidenciar que, por el momento, es la vía verde la que se ha desarrollado desde el Centro de estudios con sus plataformas de acceso inspiradas y diseñadas específicamente para actores de la comunidad educativa investigadoras e investigadores (con acceso libre de pago), y poniendo a disposición datos referentes a información estadística y documentos de estudios finales que están en un proceso de migración a una vía dorada de los datos abiertos.

Seguir avanzando hacia el desarrollo de una vía dorada parece relevante debido a varias razones. Por un lado, y como ya se ha mencionado, con la reutilización de información primaria existe una oportunidad de disminuir la cantidad de procesos de recolección de información en los establecimientos escolares del país, y reducir costos económicos y de tiempo tanto de quienes investigan como de las distintas comunidades que participan en estos procesos. Y por otro, permite avanzar en la implementación de un gobierno digital que utilice más, de mejor manera y por medio de la interacción y colaboración con la ciudadanía, evidencia para la toma de decisiones de las políticas educativas que aspiran mejorar la calidad de vida de las personas.

8. Estándares para el registro y almacenamiento de datos con potencial investigativo en el Sistema de Aseguramiento de la Calidad de la Educación (SAC)

(Versión 1.0 – agosto 2018)

El liberar información o hacerla disponible no basta para reconocer los datos automáticamente como abiertos, es necesario que estos cumplan con estándares que permitan considerarlos como tal. En otras palabras, que los datos tengan el potencial de ser accesibles, localizables, interoperables y reutilizables. Con este objetivo, el Centro de Estudios del Ministerio de Educación ha diseñado el presente documento de orientaciones para el registro y almacenamiento de la información cuantitativa y cualitativa que busca ser aplicable tanto para las investigaciones desarrolladas de manera interna en el Sistema de Aseguramiento de la Calidad como para aquellas elaboradas externamente (vía licitaciones).

Al respecto, y siguiendo los principios de acceso, localización, interoperabilidad y reutilización de la información, se presenta a continuación un conjunto de orientaciones para el registro y almacenamiento de datos generados en el marco de investigaciones cuantitativas y cualitativas.

1. Registro de los datos

Esta sección aborda los contenidos mínimos y la forma en que deben ser registrados los datos. Un adecuado registro de los datos permitirá que los datos producidos sean interoperables, es decir, que puedan ser combinados o complementados con otros, lo que en términos concretos se traduce en registrar información común y mínima en todas las investigaciones. Un riguroso registro también permitirá que sean potencialmente reutilizables, esto es, que puedan ser empleados con otros métodos, en otros momentos y por usuarios distintos.

➤ **Contenidos mínimos**

Refiere a los contenidos necesarios para completar las fichas técnicas que resumen los aspectos principales de la investigación, con el objetivo de asegurar una mejor contextualización de los análisis y de garantizar la interoperabilidad y reutilización de los datos. Se han separado en dos tipos a) los que dan cuenta del contexto de realización del levantamiento, y b) los que refieren a la unidad de observación.

a. Contexto de realización

Para una reutilización de la información es necesario registrar el contexto de desarrollo de cada proceso de levantamiento de datos que permita analizar posteriormente la información de manera contextualizada. Este proceso incluye el registro de:

- Día
- Hora
- Duración
- Lugar (refiere al espacio físico, por ejemplo, sala de clases, sala de Reuniones de hotel, hogar, entre otros)
- Ciudad y país
- Incidencias del proceso

b. Registrar las características del o los sujeto(s) objeto(s) de estudio

Junto con resguardar el marco ético de trabajo de quienes participan del estudio, se deben registrar las características que permitan contextualizar los eventuales futuros análisis.

En el caso que la unidad de observación sea(n) persona(s) natural(es), se debe registrar la siguiente información:

- RUT⁷
- Sexo
- Edad o tramo de edad
- Lugar de residencia (Ciudad/País)
- RBD⁸
- Otros criterios relevantes para la selección

En el caso que la unidad de observación o registro es un establecimiento escolar (o la captura de información se llevó a cabo en este), registrar en la base de datos la siguiente información:

- RBD⁹
- Región
- Comuna
- Área geográfica (Rural o Urbano)
- Dependencia administrativa
- Niveles educativos impartidos y tipo de formación diferenciada
- Si es en aula registrar el grado (por ejemplo, 1º básico, 4º medio, etc.)
- Otros criterios que hayan sido ocupados para su selección

⁷ En el caso que la investigación así lo amerite y se disponga de autorización de la persona. Posteriormente, el Ministerio de Educación se encargará de enmascarar los datos para el resguardo del anonimato y la protección de los datos personales, cuyo resguardo está consignado en la Ley N° 19.628 sobre Protección de Datos de Carácter Personal.

⁸ En el caso que se requiera mantener el anonimato del lugar donde estudia o trabaja la persona, estos datos serán enmascarados o eliminados posteriormente por el Ministerio de Educación.

⁹ En el caso que la investigación incluya el anonimato del establecimiento, estos datos serán enmascarados o eliminados por el Ministerio de Educación.

➤ **Formatos de registro**

Para alcanzar un estado avanzado en datos abiertos, no solo es necesario disponer de contenidos mínimos de información sino que sean registrados en formatos equivalentes y que faciliten su utilización e interoperabilidad.

a. Formatos Ficha técnica

La Ficha técnica es un documento que acompaña al conjunto de materiales de un estudio o investigación que busca identificar, de manera resumida, los objetivos, la metodología y los datos básicos de ejecución. Concretamente, este documento incluye:

Información de la investigación:

- Nombre y resumen de la investigación y/o estudio
- Objetivo general y objetivos específicos
- Metodología¹⁰ (unidades de observación, muestra y criterios de selección)
- Fechas de realización del trabajo de campo completo
- Productos asociados a la investigación

b. Formato Bases de datos (solo para investigaciones con componente cuantitativo)

Las bases de datos deben poseer una estructura en la que cada variable sea una columna y cada observación, una fila. No se deben incluir totales, subtotales, ni agrupamientos de datos. Además, se debe agregar las variables que registran las características contextuales de las unidades de análisis.

Nombre variable	Tipo	Descripción	Anchura	Decimales	Descripción	Valores permitidos

Respecto al nombre de las columnas:

- Las columnas deben tener nombres claros, pero que no sobrepasen los 50 caracteres.
- Deben nombrarse en mayúscula, separados por un guion bajo (por ejemplo: COD_DEPE, APR_IND_07).
- No deben contener espacios en blanco.
- En el caso de las columnas que presentan nombres y apellidos de personas naturales se debe identificar una columna con apellidos (paterno y luego materno) y una segunda columna, con nombres.
- En el caso de variables que sean empleadas en las bases oficiales del Ministerio de Educación, ocupar aquellos formatos. Por ejemplo:

¹⁰ La metodología debe ser lo más detallada posible, de manera que se pueda replicar el proceso en caso de que este se quiera hacer nuevamente. Se deben especificar las técnicas de recolección e instrumentos empleados.

Nombre variable	Anchura	Decimales	Descripción/Valores
AGNO	4	0	Año de la información
RBD	5	0	Rol Base de Datos del establecimiento
COD_REG_RBD	2	0	Código de región en que se ubica el establecimiento
COD_PRO_RBD	3	0	Código de la provincia en que se ubica el establecimiento
COD_COM_RBD	5	0	Código oficial3 de comuna en que se ubica el establecimiento
COD_DEPE	1	0	Dependencia administrativa:
RURAL_RBD	1	0	Corporación Municipal

c. Formato Esquema de registro (solo para investigaciones con componente cuantitativo que incluye bases de datos).

El Esquema de registro es un documento que entrega más información respecto de los archivos de datos. Este debe incluir:

- Nombre base de datos
- Año
- Responsable de la construcción del archivo: persona natural o jurídica encargada de efectuar el levantamiento de la información.
- Responsable de la recolección de la información: persona natural o jurídica encargada de realizar el levantamiento de la información.
- Unidad de observación: unidad mínima de observación.
- Número de variables.
- Número de observaciones: cantidad de unidades.

d. Formato Ficha resumen datos cualitativos (solo para investigaciones con este componente).

Para cada unidad de captura (entrevista, grupo de discusión, jornada, observación, etc.) incorporar los siguientes puntos:

a) Contexto de realización

b) Características del (los) sujeto(s) objeto(s) de estudio

c) Tipo de transcripción aplicada:

- Transcripción simple: las únicas reglas corresponden a las gramaticales y ortográficas.
- Transcripción intermedia: además de las reglas gramaticales también se incorporan referencias simples de lenguaje no verbal usualmente entre paréntesis.
- Transcripciones complejas: se incluye simbología asociada al lenguaje no verbal o contextual.

e. Formato de transcripción (solo para investigaciones con este componente)

Esta sección aplica para aquellos procesos de recolección de información que requieran el traspaso de información desde un formato audio o video a formato texto. Ya sea entrevistas, grupos focales u otras técnicas conversacionales.

La transcripción debe contener:

- En el encabezado del documento agregar los datos básicos de la transcripción: año, tipo de técnica, identificador de la transcripción, sujetos participantes y, en línea inferior, el nombre de la investigación o estudio. Todo en formato centrado. Ejemplo:

2016 - Entrevista 01 - Docentes
Estudio Recursos Escolares

- En la primera parte, señalar las abreviaturas consideradas para identificar a los participantes (respetando el anonimato).
- Para las intervenciones de quienes moderan o entrevistan, emplear letras mayúsculas.
- Asignar una línea entre una intervención y otra.

2. Almacenamiento

Regular el almacenamiento de los datos también constituye un elemento fundamental para avanzar en una mejor implementación de datos abiertos pues permite mejorar el acceso, especialmente en términos de formato, y por medio de la normalización otorga las condiciones necesarias para la reutilización de la información.

Todo el material deberá estar en una única carpeta. El nombre de la carpeta seguirá el siguiente formato: [año] [-] [nombre resumido de investigación].

Ejemplos:

2017- Convivencia escolar

2016- Recursos educativos

2015- Educación artística

Los archivos deben ser registrados según los lineamientos recomendados por las Normas Técnicas de Publicación de Datos Abiertos Chile; esto es, Lower Camel Case “las palabras se escriben juntas y sin espacio, donde la primera letra de la primera palabra se escribe con minúscula, y las primeras letras de las subsiguientes palabras se escriben con mayúscula” (Gobierno Abierto, 2013).

La carpeta deberá incluir los siguientes archivos con los siguientes formatos de nombre:

Tabla 1. Archivos carpeta informe final

Archivo	Formato nombre	Ejemplos	Formato archivo
Informe final	InformeFinal[nombre resumido de investigación][Añofinalización]	InformeFinalConvivenciaEscolar2014 InformeFinalPIE2012	*.doc (o *.docx) *.pdf
Informes avances	InformeAvance[n°informe de avance][nombre resumido de investigación] [Añofinalización]	InformeAvance1ConvivenciaEscolar2014 InformeAvance2 ConvivenciaEscolar2014 InformeAvance1PIE2012 InformeAvance2PIE2012	*.doc (o *.docx) *.pdf
Ficha técnica	FichaTecnica[nombre resumido de investigación][Añofinalización]	FichaTecnicaConvivenciaEscolar2014 FichaTecnicaPIE2012	*.doc (o *.docx) *.pdf

Además, debe incluir las siguientes carpetas:

➤ Carpeta consentimientos

Debe incluir la totalidad de consentimientos y asentimientos aplicados en la investigación. En el caso de haber sido obtenidos en papel, estos deben ser digitalizados.

➤ Instrumentos

Debe incluir la totalidad de instrumentos aplicados en la investigación, es decir, pautas de entrevistas, pautas de observación, entre otras. También se debe acompañar, dependiendo del caso, de manuales de aplicación, orientaciones para capacitaciones u otro material que permita aplicar el instrumento.

➤ Datos cuantitativos

La carpeta con los datos cuantitativos deberá incluir las bases de datos y los esquemas de registros de cada una de ellas, según el siguiente formato:

Tabla 2. Archivo carpeta de datos

Archivo	Formato nombre	Ejemplos	Formato archivo mínimo
Esquemas de registros	ER[nombre resumido de investigación] [Sujeto/Objeto/Proceso de Estudio][año de finalización de investigación]	ERConvivenciaEscolarD centes2014 ERPIEProyectos2012 EROtrosIndicaresEstudia ntes2017	*.doc (o *.docx) *.pdf
Bases de datos	BD[nombre resumido de investigación] [Sujeto/Objeto/Proceso de Estudio][año de finalización de investigación]	BDConvivenciaEscolarD centes2014 BDPIEProyectos2012 BDOtrosIndicaresEstudia ntes2017	Ver tabla 3

Los datos deben ser almacenados en formatos digitales que puedan ser publicables y reutilizables, lo que se modificará dependiendo del tipo de registro, según los siguientes formatos:

Tabla 3. Formato de los archivos

Tipo de dato	Formato mínimo	Formatos complementarios
Genéricos	CSV	SDMX, SPSS, STATA, XML
Geográficos	GeoTIFF, WFS, WCS	GML, ArcGIS, KML
Redes	NET	

Fuente: Norma Técnica para Publicación de Datos Abiertos en Chile. Adaptación y traducción propia de "Guide to depositing qualitative data at DARIS".

➤ Datos cualitativos

La carpeta con los datos cualitativos deberá incluir la ficha resumen de datos cualitativos y las unidades de observación¹¹.

¹¹ Cada unidad de observación: entrevista (E), entrevista grupal (EG) grupo focal (GF), grupo de discusión (GD), observación participante (OP), observación no participante (ONP), jornada (J), entre otras, deberá ser registrado en un archivo de datos separado.

Tabla 4. Archivos carpeta de datos cualitativos

Archivo	Formato nombre	Ejemplos	Formato archivo mínimo
Ficha resumen datos cualitativos	ResumenCualitativo[nombre resumen del estudio] [año de finalización de investigación]	ResumenCualitativoRecursosEscolares2016 ResumenCualitativoConvivencia2014 ResumenCualitativoF5326412012 ResumenCualitativoEducativaPET2012	*.doc (o *.docx) *.pdf
Unidades de observación	[sigla de técnica][número identificador] [Sujeto/Objeto de Estudio][nombre resumen del estudio] [año de finalización de investigación]	E01DocentesRecursosEscolares2016 GF03estudiantesConvivencia2014 O03AulasF5326412012 J11ComunidadEducativaPET2012	*.doc (o *.docx) *.pdf

Los datos deben ser almacenados en formatos digitales que puedan ser publicables y reutilizables, lo que se modificará dependiendo del tipo de registro.

Tabla 5. Formatos según tipo de dato

Tipo de dato	Formato mínimo	Formatos complementarios
Texto	Adobe Portable Document Format PDF (*.pdf) MS Word (*.doc, *.docx)	OpenDocument Text (*.odt) Rich Text Format RTF (*.rtf) Mark-up Language HTML (*.html) eXtensible Mark-up Language XML according to an appropriate schema (*.xml) Unicode, Non-Unicode and ASCII (*.txt)
Audio	Waveform Audio Format (*.wav)	MPEG-1 Audio Layer 3 (*.mp3) Audio Interchange File Format AIFF (*.aif, *.aiff)
Imágenes	Joint Photographic Experts Group JPEG (*.jpeg, *.jpg)	TIFF (*.tif, *.tiff) Adobe Portable Document Format PDF (*.pdf) JPEG 2000 (*.jpg2, *.jpx) PNG (*.png) GIF (*.gif) BMP (*.bmp) Drawing Interchange File Format DXF (*.dxf) PostScript (*.eps) Adobe Illustrator (*.ai) Scalable Vector Graphics SVG (*.svg)
Video	MPEG-4 (*.mpg4)	Moving Picture Experts Group MPEG-2 (*.mpg2) Motion JPEG 2000 (*.mj2) QuickTime (*.mov) Lossless AVI (*.avi)

Fuente: Adaptación y traducción propia de "Guide to depositing qualitative data at DARIS".

Bibliografía

Back, B. (2016). Governments are the Tip of the Open Data Iceberg. Recuperado de <https://blog.datazar.com/governments-are-the-tip-of-the-open-data-iceberg-1a22be783aaa>

Canales, M. (2006). Metodologías de investigación social. Introducción a los oficios. Santiago: Lom Ediciones.

Concha, G. & Naser, A. (2012). Datos abiertos: Un nuevo desafío para los gobiernos de la región. Naciones Unidas – CEPAL. Instituto Latinoamericano y del Caribe de Planificación Económica y Social (ILPES), Serie 74. Santiago, marzo de 2012.

Datos abiertos Gobierno de España (2017). Open Access: la vía verde, la vía dorada y la vía diamante. Recuperado de: <http://datos.gob.es/es/noticia/open-access-la-verde-la-dorada-y-la-diamante>

Doorn, P. & Dillo, I. (2016). FAIR data in trustworthy data repositories (en línea). Recuperado de https://www.slideshare.net/EUDAT/fair-data-in-trustworthy-data-repositories-webinar-1213-december-2016?gid=b369efe3-48d4-4442-a548-77d9be8ca9be&v=&b=&from_search=1

European Commission (2003). Agenda Digital: hacer fructificar los datos de las administraciones. Press Release Database. Recuperado de http://europa.eu/rapid/press-release_IP-11-1524_es.htm?locale=en

Ferrer-Sapena, A. (2017). Los datos en el contexto actual: utilidades y servicios. V Festival de los libros y III Jornadas del Conocimiento “El mundo de los datos”. Universidad Politécnica de Valencia, España.

Ferrer-Sapena, A. & Sánchez-Pérez, E. (2013). Open data, big data: ¿hacia dónde nos dirigimos? Anuario ThinkEPI 7, 150-156.

Ferrer-Sapena, A., Peset, F. & Aleixandre-Benavent, R. (2011). Acceso a los datos públicos y su reutilización: open data y open government. El profesional de la información 20(3), 260-269.

Gobierno abierto (2015). Carta internacional de datos abiertos. Blog Abierto. Recuperado de <http://gobabiertomx.org/carta-internacional-de-datos-abiertos/>

González, L. & Peset, F. (2013). Datos de investigación: reflexiones sobre su acceso abierto. International open access week 2013 (Conference paper).

Hofmann, A., Ramírez, A. & Bojórquez, J. (2012). La promesa del gobierno abierto. Recuperado de http://www.infodf.org.mx/comsoc/campana/2012/prom_gob_abierto.pdf

Information Commissioner's Office (2017). Artificial Intelligence, Machine Learning and Data Protection. Big Data. Recuperado de <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>

Larraín, H. & Gazmuri, J. (2005). Historia de la Ley N° 20.285, sobre acceso a la información pública. Biblioteca del Congreso Nacional de Chile/BCN. Recuperado de https://www.bcn.cl/historiadelaley/fileadmin/file_ley/6357/HLD_6357_749a0d2dec7072ac83d52ebf0f2ff393.pdf

Ley N° 20.529. Sistema Nacional de Aseguramiento de la Calidad de la Educación Parvularia, Básica y Media y su Fiscalización. Diario Oficial de la República de Chile, Santiago, 08 de junio de 2015. Recuperado de <https://www.leychile.cl/N?i=1028635&f=2016-03-01&p=>

Ley N° 20.285. Sobre acceso a la información pública. Diario Oficial de la República de Chile, Santiago, 05 de enero de 2016. Recuperado de <https://www.leychile.cl/Navegar?idNorma=276363&buscar=transparencia>

Mejía, J. (2008). Epistemología de la investigación social en América latina. Desarrollos en el siglo XXI. Cinta Moebio 31, 1-13.

Méndez, E. (año). "Cool" metadata for FAIR data (en línea). Recuperado de <https://www.slideshare.net/ResearchDataAlliance/cool-metadata-for-fair-data>

Ministerio de Educación, Centro de Estudios (2018). Evidencia para decisiones informadas, 2014-2017. Santiago: Autor.

Gobierno de Chile, Unidad de Modernización y Gobierno Digital, Minsegespres (2013). Norma Técnica para Publicación de Datos Abiertos en Chile.

Gobierno de Chile, Minsegespres (2016). Estrategia Chile Datos Abiertos. United Nations Public Administration Network. Recuperado de <http://workspace.unpan.org/sites/Internet/Documents/UNPAN96571.pdf>

Nina-Alcocer, V., Blasco-Gil, Y. y Peset, F. (2013). Datasharing: guía práctica para compartir datos de investigación. El profesional de la información 22,6, 562-568.

Obama, B. (21 de enero de 2009) Transparency and Open Government. Memorandum for the heads of executive departments and agencies. The White House, Washington. Recuperado de <https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government>

Organización para la Cooperación y el Desarrollo Económicos (2017). Building a data-driven public sector in Norway. En Digital Government Review of Norway (155-207). Paris: OECD Publishing.

Organización para la Cooperación y el Desarrollo Económicos (2017). Government at a Glance 2017. Paris: OECD Publishing.

Open Data Charter (2015). Principios carta internacional de datos abiertos. Recuperado de <https://opendatacharter.net/principles-es/>

Open Data Institute (2017). The Data Ethics Canvas user guide. Recuperado de <https://theodi.org/article/data-ethics-canvas/>

Piñera, S. (2012). Imparte instrucciones sobre Gobierno Abierto. Gabinete presidencia N° 005. Santiago: Gobierno de Chile.

Peñaloza, A. & Osorio, M. (2005). Curso-Taller: Elaboración de instrumentos de Investigación. Recuperado de: <https://es.scribd.com/document/360399542/guia-para-elaboracion-de-instrumentos-pdf>.

Peset, F., Alexandre-Benavent, R., Blasco-Gil, Y., & Ferrer-Sapena, A. (2017). Datos abiertos de investigación. Camino recorrido y cuestiones pendientes. Anales de Documentación20(1).

Rodríguez, P., González-Zapata, F., Inostroza, J. Roeschmann, J. & Ledermann, R. (2017). 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. Documento con líneas de estudio/trabajo para el programa de modernización del sector público. Chile: Dirección de Presupuestos, Ministerio de Hacienda.

Rodríguez, P., González-Zapata, F., Inostroza, J. Roeschmann, J. & Ledermann, R. (2017). 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. Chile: Dirección de Presupuestos, Ministerio de Hacienda.

Solana, A. (12 de febrero de 2011). Liberando datos para construir ciudades más inteligentes. La vanguardia. Recuperado de <http://www.lavanguardia.com/internet/20110212/54114005322/liberando-datos-para-construir-ciudades-mas-inteligentes.html>

The Royal Society (2012). Science as an open enterprise. London: The Royal Society Science Policy Centre report.

Trias, J. (2013). Uso de datos administrativos para la evaluación de impacto. Banco Mundial. Taller Internacional de Encuestas y Evaluación de Impacto en políticas públicas. Recuperado de <http://pubdocs.worldbank.org/en/441021463422864349/Core-Sesion-1c-Usode-datos-administrativos-J-Trias.pdf>

UK Statistics Authority (2018). National Statistician's Data Ethics Advisory Committee. Recuperado de <https://www.statisticsauthority.gov.uk/about-the-authority/committees/nsdec/>

Web Foundation (2013). Open Data Barometer, 2013 Global Report. World Wide Web Foundation. Recuperado de <http://opendatabarometer.org>

Web Foundation (2017). Open Data Barometer, 2017 Global Report. World Wide Web Foundation. Recuperado de <http://opendatabarometer.org>

Anexos

Anexo 1. Ministerios y páginas web revisadas

Ministerio	Sitios web
Interior y Seguridad Pública	http://cead.spd.gov.cl/estadisticas-delictuales/ http://www.subdere.gov.cl/documentacion
Relaciones Exteriores	https://minrel.gob.cl/
Defensa Nacional	https://www.anepe.cl/centro-de-estudios-estrategicos-2/
Hacienda	http://www.hacienda.cl/documentos http://modernizacion.hacienda.cl/estudios
Secretaría General de la Presidencia	http://www.minsejpres.gob.cl/documentos/
Secretaría General de Gobierno	http://www.msgg.gob.cl/wp/index.php/listado-de-estudios/
Economía, Fomento y Turismo	http://www.economia.gob.cl/estudios-y-encuestas
Desarrollo Social	http://siist.ministeriodesarrollosocial.gob.cl/ http://observatorio.ministeriodesarrollosocial.gob.cl/indicadores/
Educación	http://cedoc.mineduc.cl/mvc/buscador/buscarDocumentoAvanzado http://datosabiertos.mineduc.cl/ https://centroestudios.mineduc.cl/
Justicia y Derechos Humanos	http://www.minjusticia.gob.cl/documentos/
Trabajo y Previsión Social	http://www.sil.mintrab.gob.cl/ http://www.mintrab.gob.cl/centro-de-estudios/
Obras Públicas	http://www.mop.cl/CentrodeDocumentacion/Paginas/InvestigacionesyEstudios.aspx
Salud	http://www.bibliotecaminsal.cl/ http://www.deis.cl/
Vivienda y Urbanismo	http://www.minvu.cl/opensite_20061113165233.aspx http://www.minvu.cl/opensite_20070817104515.aspx http://www.observatoriohabitacional.cl
Agricultura	https://www.minagri.gob.cl/centro-de-documentacion/
Minería	http://www.minmineria.gob.cl/informacion-estadistica/
Transporte y Telecomunicaciones	http://www.mtt.gob.cl/estudiosyestadisticas http://www.subtel.gob.cl/estudios-y-estadisticas/
Bienes Nacionales	http://www.bienesnacionales.cl/ www.ide.cl
Energía	http://energiaabierta.cne.cl/estudios/
Medio Ambiente	http://sinia.mma.gob.cl/
Deporte	http://www.mindep.cl/estudios-ministerio-del-deporte/
Mujer y la Equidad de Género	https://www.minmujeryeg.cl/ministerio/estudios-y-capacitacion/
Culturas, las Artes y el Patrimonio	http://www.cultura.gob.cl/ministerio/

Anexo 2. Instrumento de revisión sitios web gubernamentales

Nombre del ministerio	
¿Posee sitios web de estudios y/o estadísticas?	0: No 1: Sí
¿Sitio único de acceso a datos o información del ministerio?	0: No 1: Sí
¿Existe un catálogo para la búsqueda de datos?	0: No 1: Sí
¿Poseen datos abiertos (sin registro de acceso)?	0: Todos los casos 1: Algunos casos 2: Ningún dato
¿Dispone de información oficial del servicio?	0: No 1: Sí
¿Dispone de información procedente de otros servicios?	0: No 1: Sí
¿Existen informes o reportes de investigación disponibles para la comunidad?	0: No 1: Sí
¿Cuál es el último año de los informes o reportes de investigación disponibles?	
¿Está estandarizado el modo para disponibilizar informes?	0: No 1: Sí
¿Los informes se encuentran en más de un formato?	0: Todos los casos 1: Algunos casos 2: Ningún dato
¿Existen informes en formato Word?	0: No 1: Sí
¿Existen informes en formato PDF?	0: No 1: Sí
¿Existen informes en formato PPT?	0: No 1: Sí
¿Existen informes en otro formato?	0: No 1: Sí
¿Cuál otro formato de informes?	
¿Los informes disponibilizan contenidos asociados (subproductos) a los informes de investigación?	0: Todos los casos 1: Algunos casos 2: Ningún dato
¿Están estandarizados los contenidos asociados a los informes de investigación?	0: No 1: Sí
¿Existen bases de datos asociadas a los informes de investigación?	0: No 1: Sí

¿Existen transcripciones asociadas a los informes de investigación?	0: No 1: Sí
¿Existen instrumentos asociadas a los informes de investigación?	0: No 1: Sí
¿Existen insumos administrativos asociados a los informes de investigación?	0: No 1: Sí
¿Existen informes parciales asociados a los informes de investigación?	0: No 1: Sí
¿Hay disponibles bases de datos?	0: No 1: Sí
¿Las bases de datos son de origen administrativas?	0: No 1: Sí
¿Las bases de datos son producto de investigaciones?	0: No 1: Sí
¿Las bases de datos son de un origen externo al servicio?	0: No 1: Sí
¿Las bases de datos se encuentran estandarizadas?	0: No 1: Sí
¿Las bases son interoperables?	0: No 1: Sí
¿Todas las bases de datos se encuentran disponibles en más de un formato?	0: No 1: Sí
¿Existen informes en bases de datos en formato XLS o XLSX?	0: No 1: Sí
¿Existen informes en bases de datos en formato CVS?	0: No 1: Sí
¿Existen informes en bases de datos en formato JSON?	0: No 1: Sí
¿Existen informes en bases de datos en formato HTML?	0: No 1: Sí
¿Existen informes en bases de datos en formato SAV?	0: No 1: Sí
¿Existen informes en bases de datos en formato STATA?	0: No 1: Sí
¿Cuál es el último año de bases de datos disponibles?	
¿Existe sistema visualización de datos?	0: No 1: Sí
¿Existen disponibles datos cualitativos?	0: No 1: Sí

¿Existen datos cualitativos en formato Word?	0: No 1: Sí
¿Existen datos cualitativos en formato PDF?	0: No 1: Sí
¿Existen datos cualitativos en formato MP3?	0: No 1: Sí
¿Existen datos cualitativos en formato MP4?	0: No 1: Sí
¿Existen datos cualitativos en otro formato?	0: No 1: Sí
Comentarios adicionales	