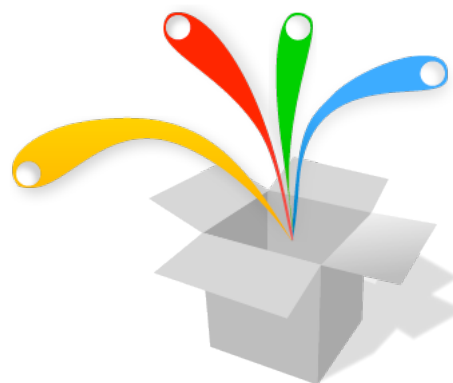


Propuesta de Norma Técnica para Publicación de Datos en Chile

*Bases para la elaboración y discusión de una norma
para la publicación de datos públicos en Chile*

Enero de 2012



Índice de Contenidos

Índice de Contenidos

1. Introducción

2. Propósito

3. Condiciones Previas

4. Preparación, Validación y Formatos

4.1 Formatos

4.1.1 Datos genéricos

4.1.2 Datos geográficos

4.1.3 Imágenes

4.1.4 Datos estadísticos

5. Publicación

5.1 Metadatos

5.2 Catálogos

5.3 Publicación en data.gob.cl

5.4 Licencias de publicación de datos

6. Coordinación y Actualización

6.1 Hitos y Plazos

6.2 Condiciones Especiales

6.3 Frecuencia de Actualización

Anexo 1 - Principios y Consideraciones

Anexo 2 - Referencias Normativas

Anexo 3 - Términos y Definiciones

Anexo 4 - Abreviaturas y Acrónimos

Anexo 5 - Referencias Bibliográficas

1. Introducción

Open Data (Datos Abiertos) es un concepto y una iniciativa ya instalada en diversos países y que persigue la publicación de conjuntos de datos, que las instituciones y en particular los Estados tienen en su poder, con el fin de que éstos sean divulgados y potencialmente utilizados por la Sociedad en general y en particular por entidades que puedan agregar valor a dichos datos y con ello elaborar productos de beneficio para comunidades de ciudadanos. Existe una gran variedad de datos que califican para estos fines: geográficos, meteorológicos, de tráfico, de la gestión del gobierno, el uso de los recursos fiscales, entre otros.

Los comienzos de esta iniciativa se atribuyen al gobierno británico con su iniciativa *Opening up Government*¹ (2010). La Unión Europea incluso antes, emanó la *Directiva 2003/98/CE*², relativa a la reutilización de la información del sector público (2003). A la fecha, el Reino Unido y Estados Unidos son quienes llevan la delantera en este campo, sin embargo, existen iniciativas en este ámbito en varios países.

En este contexto surge el proyecto Datos.gob.cl, iniciativa del Gobierno de Chile dirigida a proveer al público de un acceso expedito, abierto, y sin restricciones de uso, a conjuntos de datos gubernamentales, en formatos estructurados, que permitan su reutilización tanto por terceros como de manera automatizada, en el entendimiento que esto puede significar un importante estímulo a entidades que pueden agregar valor a dicha información y también eventualmente potenciar actividades de emprendimiento asociadas a la divulgación y agregación de información con relevancia para el ciudadano.

2. Propósito

En un modelo de datos abiertos de gobierno, las distintas instituciones de la Administración del Estado ponen a disposición de los ciudadanos e instituciones públicas o privadas, los datos en su poder.

Por este motivo, un aspecto de gran relevancia al abordar la preparación y exposición de datos públicos en formato electrónico en Chile es normalizar dichos procesos, a fin de que cada institución utilice metodologías, procedimientos y prácticas comunes que garanticen la organización lógica, física y de formatos de los datos publicados.

¹ Ver <http://data.gov.uk/>

² Ver http://www.aporta.es/c/document_library/get_file?uuid=824b77a0-b033-4c9d-b8a6-52807ea30648&groupId=10128

En consecuencia, el objetivo del presente documento es establecer los lineamientos que se deben respetar para facilitar el acceso y uso de la información pública, a través del establecimiento de las características mínimas que deben cumplir los conjuntos de datos así como el proceso de publicación, entendiendo por éste el proceso de hacer exposición pública permanente de conjuntos de datos por parte de una institución, para el uso abierto de todos los ciudadanos, ya sea por medio de interfaces humanas o procedimientos automatizados. Incluye la publicación de los datos propiamente tal, así como sus correspondientes metadatos.

Los datos considerados en esta norma son aquellos contenidos y registrados en cualquier formato y que son recogidos, producidos y/o recibidos por los órganos de Administración del Estado durante el ejercicio de sus actividades o de acuerdo a funciones encomendadas, los cuales son mantenidos, organizados y conservados como objeto de antecedente o consulta para la propia institución, la Administración Pública y la ciudadanía en general.

3. Condiciones Previas

Toda institución interesada en publicar sus ‘datasets’ o “conjuntos de datos”³, deberá mantener un repositorio de datos públicos, asegurando las condiciones operacionales para garantizar la continuidad de los servicios (servicio 24x7, condiciones mínimas para servicio a múltiples usuarios simultáneos, interfaz de acceso unificada). Adicionalmente, debe contar con procedimientos y dirección centralizada en cada organización y tener un responsable de su organización, mantenimiento y funcionamiento, que asegure la integridad, autenticidad y permanencia en el tiempo de los datos que posee. Al respecto se sugieren las siguientes consideraciones:

- (a) **Control de acceso:** solamente las personas o sistemas autorizados deben tener acceso a la publicación y modificación de datos y metadatos. Se recomienda documentar la política de acceso, y usar mecanismos básicos de autorización (usuario, contraseña).
- (b) **Integridad:** Debe asegurarse que los datos no sean adulterados durante la transferencia. Se recomienda usar URL fija como testigo o firma, y en un nivel más exigente, usar SSL.

³ Estos términos se usarán indistintamente a lo largo de este documento.

- (c) **Autenticidad:** Debe asegurarse que los datos provienen realmente de la fuente de su publicador. Se recomienda usar URL fija como testigo o firma, y en un nivel más exigente, usar SSL.

Por otra parte, cada organismo encargado de la administración del repositorio será el responsable de la integridad, autenticidad y del sistema de licenciamiento de los datos públicos que posee su repositorio. Opcionalmente podrá tener un sistema de certificación de datos.

4. Preparación, Validación y Formatos

Con el fin de facilitar el acceso expedito, abierto, y sin restricciones de uso de conjuntos de datos gubernamentales, antes de la publicación, el responsable institucional debe procurar que los datasets cumplan con las siguientes condiciones generales:

1. Los datos a ser publicados deben estar en la forma menos procesada posible, es decir, previos a cualquier agregación estadística o pre-procesamiento.
2. Los datasets deben estar en formatos abiertos, no propietarios, estables y de amplio uso.
3. Se debe considerar la disponibilidad del dataset en múltiples formatos.
4. Cada dataset debe tener un identificador único y permanente, según el estándar URI..
5. Para la denominación de los archivos se recomienda el estilo *lowerCamelCase*. Ejemplo: *lluviasSeptiembre2010.xml*
6. Cada dataset debe contar con la información sobre sus datos y de sus metadatos. Debe ser posible recuperar el sentido de la codificación de datos usada y el significado de los datos.
7. Los metadatos a utilizar son los definidos en esta norma (*ver capítulo 7.1*).
6. Para datasets de gran tamaño, se recomienda dividirlo en datasets más pequeños que permitan una más sencilla manipulación.
8. En los casos de que los dataset estén constituidos por uno o más archivos, cada dataset completo debe ser posible de reconstituir a partir de los metadatos que se describen en la distribución del dataset.
9. En el caso de datasets derivados, entendiendo por éstos los dataset que están esencialmente formados a partir de otros datasets ya publicados, considerar que:
 - a. aquellos que sean parte o subconjunto de otro dataset, deben indicar su procedencia esto es, el dataset del cual se tomó los datos que lo conforman. Esta información debe estar contenida en los metadatos del nuevo dataset.

- b. aquellos que son integración completa o parcial de varios datasets, deben indicar su procedencia, esto es, los datasets de los cuales se tomaron los datos que lo conforman. Esta información debe estar en los metadatos del nuevo dataset.
 - c. aquellos que son producto del procesamiento de otros datasets, deben indicar la procedencia de donde se tomaron los datos que lo conforman.
10. Para datasets con validez temporal, se debe indicar desde cuándo y hasta cuándo son válidos.
 11. Para datasets con actualización periódica, se debe indicar la periodicidad de actualización.
 12. Se recomienda que los metadatos para coberturas y datos georeferenciados contengan los campos de metadatos obligatorios de la norma ISO 95139 usados por el Servicio Nacional de Información Territorial (SNIT).
 13. Para datos estadísticos se debe incluir los metadatos correspondientes a metodología y calidad de datos.

Particularmente con los datos de tipo estadístico, se recomienda especial cuidado en su preparación ya que cada publicador es responsable de velar por el secreto estadístico de los datos publicados. Considerar al menos:

- a. Se debe evitar publicar relaciones entre datos que son privados que permitan a terceros obtener información personal privada.
- b. Son particularmente críticos los *identificadores* personales, es decir aquellos datos que permiten la identificación única de una persona. Particularmente sensibles son: Números de identificación de documentos personales (RUT, carnet manejar, trabajo, salud, etc.); Identificación geográfica (dirección, casilla correo, etc.); e Identificación virtual (email, teléfono, etc.).

4.1 Formatos

Sobre los formatos de los datasets, la presente norma agrupa cuatro grandes categorías, una genérica y tres especializadas, cuyas características técnicas exigen un tratamiento particular.

4.1.1 Datos genéricos

Datos que no tienen aplicaciones especializadas, y que normalmente se corresponden con datos provenientes de bases de datos y reportes como planillas de cálculo y cuadros de información.

1. Características de formatos a usar:

- (a) Se debe considerar el uso de formatos estándares, abiertos (no propietarios). En casos muy justificados (datos históricos, transformaciones costosas), usar formatos propietarios o aún no estandarizados, pero indicando el software, versión y productor que permite procesar esos formatos.
 - (b) Para datos de planillas y bases de datos tradicionales, se debe considerar el uso de formatos CSV y XML. Se recomienda limitar el uso de HTML o XHTML para datasets pequeños y en caso de usar éstos últimos, se recomienda incorporar RDFa.
2. Protocolos de pre-procesamiento para facilitar el intercambio, la organización y el procesamiento de los archivos:
- (a) Compresión de grandes archivos: Se debe considerar usar protocolos de compresión solamente para archivos muy grandes, y altamente compresibles. En ese caso, se recomienda usar protocolos sin pérdida, y abiertos (no propietarios). Ejemplo: gzip, bzip2, ppmdi, etc.
 - (b) Empaquetamiento de archivos: Se debe considerar protocolos abiertos y estándares. Ejemplo: tar. Se debe indicar explícitamente en los metadatos el protocolo usado.

4.1.2 Datos geográficos

Datos que provienen de, o son de uso en aplicaciones geográficas. Están usualmente en formatos específicos para representar la superficie terrestre o la geometría de objetos en el espacio geográfico.

1. Los formatos de publicación de coberturas y datos georeferenciados deberán estar en concordancia con los formatos definidos por el proyecto de normas de manejo de información geográfica de SNIT.
2. Como recomendación de base, los formatos de publicación pueden variar dependiendo del tipo de uso que tengan los datos, distinguiendo entre formatos orientados a los objetos (vectoriales) o orientados a la superficie (raster o imagen).
 - (a) Para visualización de imágenes, los formatos recomendados son JPEG, PNG, cuyo tamaño de píxeles depende de la escala de representación. Para uso en servidores de mapas y en análisis, el formato GeoTIFF representa imágenes TIFF cuyo encabezado incluye la extensión geográfica de la imagen.
 - (b) Para formatos vectoriales y de visualización en la Web, SVG (scalable vector graphics) es un formato no propietario de uso masivo para exportación desde

servidores de mapas. Para servidores de mapas y análisis, se recomienda el formato GML no propietario del OpenGIS Consortium.

Hay algunos formatos propietarios de uso masivo que son aceptables: shapefile es un formato propietario (ARGIS) de uso masivo que puede ser importando por la mayoría de los sistemas actuales. Contiene un archivo .shp para las geometrías y un .dbf para datos georeferenciados a estas geometrías; KML de Google es una implementación del GML que tiene amplio uso para despliegue sobre Google Earth/Map.

3. Como sistema de referencia para coordenadas se recomienda usar WGS84/SIRGAS 2000.

4.1.3 Imágenes

Datos que provienen esencialmente de fotografías.

1. Se debe usar tipos de archivos de imágenes basados en el método JPEG de compresión con pérdida, usualmente con extensión “.jpg”, y los archivos basados en el método PNG (Portable Network Graphics) de compresión sin pérdida, usualmente con extensión “.png”.
2. Para formatos vectoriales y de visualización en la Web que necesitan renderización, se recomienda SVG (scalable vector graphics).

4.1.4 Datos estadísticos

Datos usados por usuarios especializados del área estadística. Están en formatos particulares para reuso y explotación de aplicaciones estadísticas especializadas.

Como formato de datos se recomienda el Statistical Data and Metadata Exchange (SDMX) (estándar OCDE). Por su amplio uso son aceptables: SPSS, STATA (en uso hoy día). En este último caso se debe considerar además publicar una distribución de los datos en formato CSV.

Para más antecedentes de este tipo de datos, el organismo encargado de las estadísticas y censos oficiales de La República es el Instituto Nacional de Estadísticas (Ver anexo 2)

5. Publicación

Corresponde al proceso de hacer exposición pública permanente de los datasets o conjuntos de datos por parte de una institución una vez preparados para el uso abierto de todos los ciudadanos. Incluye la publicación de los datos propiamente tal, así como sus correspondientes metadatos.

5.1 Metadatos

Los metadatos son una herramienta fundamental para organizar, clasificar, relacionar y razonar sobre los datos. Es importante recordar que la cantidad y calidad de metadatos asociados a un dataset es un compromiso: más y mejores metadatos permiten una mejor clasificación y una mejor búsqueda y recuperación de los recursos. Pero, es importante hacer notar que ello mismo compromete la facilidad y simpleza para publicar datasets. Los publicadores deben considerar este compromiso.

Por ello esta norma define un conjunto mínimo de metadatos obligatorios, y un conjunto recomendado más amplio.

Metadatos Mínimos

Los metadatos obligatorios que debe incluir la publicación de un dataset son:

1. **Título:** nombre que se le da al dataset.
2. **Descripción:** un resumen breve en lenguaje natural de los contenidos del dataset.
3. **Publicador:** nombre de la institución o entidad responsable de hacer el dataset públicamente disponible.
4. **Categoría:** clasificación conceptual básica del dataset en sistemas de categorías o taxonomías disponibles.
5. **Fecha de Publicación:** punto del tiempo asociado al evento de publicación del dataset.
6. **Distribución:** medio y características físicas de una particular presentación del dataset. Un mismo dataset puede tener varias distribuciones (por ej.: en formato XML, en formato comprimido, en formato CSV, etc.).

A su vez, cada distribución debe contener los siguientes metadatos:

- (a) **Identificador:** URL[4] del archivo.
- (b) **Tamaño:** volumen físico utilizado por el archivo, en unidades estándares.

- (c) *Formato*: modo o medio principal de representación de los datos, normalmente identificados en la extensión del nombre del archivo.
- (d) *Descripción*: características adicionales (por ej.: tipo de compresión, versión de algún estándar, etc.) necesarias de conocer para su procesamiento.
- (e) *Fecha*: fecha de publicación de esta distribución del archivo.

Metadatos Recomendados

Los metadatos opcionales que es deseable que incluya la publicación de un dataset son:

1. *Etiquetas*: términos clave para describir el dataset.
2. *Autoría/Crédito*: entidad o persona primariamente responsable de la producción del dataset.
3. *Licencia*: información sobre los derechos de propiedad y de uso del dataset.
4. *Cobertura*:
 - (a) *Ubicación geográfica*: puntos o zonas geográficas o espaciales asociadas al dataset.
 - (b) *Cobertura temporal*: puntos o períodos de tiempo asociados al dataset.
 - (c) *Granularidad*: nivel de detalle con que se presentan los datos.
5. *Frecuencia de Actualización*: frecuencia temporal con que se actualiza el dataset.
6. *Referencias*: relación con otros datasets. Enlaces a otros conjuntos de datos. Linaje del dataset en el caso de datos derivados o integrados.
7. *Información asociada*: otras dimensiones de la descripción del dataset.
 - (a) *Metodología*: metodología usada en la recolección de los datos y en la producción del dataset.
 - (b) *Calidad de los datos*: incluye antecedentes de diseño, muestra, error y herramientas de análisis usadas para generar el dataset.
 - (c) *Diccionario*: descripción del esquema y tipos de datos usados en el dataset.

5.2 Catálogos

Un catálogo es la entrada organizada dirigida al usuario de los datasets publicados por el organismo, dirigido a facilita el acceso simple, abierto, y sin restricciones de uso de los datos que posee. El catálogo puede considerarse como la organización de los metadatos de los datos del repositorio, más metadatos y enlaces a otros datos relacionados relevantes para los usuarios.

A nivel de Gobierno de Chile habrá un catálogo central. Cada organización puede poseer uno propio, el que debe estar alineado con el catálogo central, es decir, tener entradas en el catálogo central para cada uno de los datasets que posea.

1. El catálogo debe contener un listado exhaustivo de los datasets presentes en el repositorio (es decir, todo dato publicado en el repositorio debe poder ser accesible a través del catálogo).
2. El catálogo debe permitir la navegación, búsqueda y consulta simple de los datos contenidos en el repositorio. Para esto implementará todas las funcionalidades necesarias que se requieran (categorizaciones, navegación, búsqueda de texto, lenguaje de consulta, etc.).
3. En una etapa de desarrollo avanzado el catálogo debiera proveer posibilidades de acceso a agentes automatizados para buscar y recuperar datos.
4. Los catálogos utilizarán un conjunto de metadatos mínimos definidos en esta norma, y los extenderán con estándares de uso internacional o nacional.
5. El catálogo debe proveer funcionalidades de retroalimentación de los usuarios, así como consulta y sugerencias.
6. Cada catálogo local debe estar enlazado con el catálogo central.
7. El responsable del Catálogo debe velar por su funcionamiento, mantener su disponibilidad y analizar sus estadísticas de uso.

5.3 Publicación en datos.gob.cl

La Unidad de Modernización y Gobierno Electrónico de MINSEGPRES, proveerá la conectividad y mantención del servicio web datos.gob.cl donde se referenciarán los datos públicos abiertos del Gobierno de Chile (Ver 6.1 sobre Hitos y Plazos). Habrá dos modalidades para la utilización de este recurso:

1. Los conjuntos de datos se remiten a la Unidad de Modernización y Gobierno Electrónico para ser publicados en la url citada, o bien,
2. Las instituciones del Estado proveen el servicio de publicación de sus propios datos y éstos sólo se referencian hacia el sitio citado.

5.4 Licencias de publicación de datos

La utilización de licencias es esencial para dar claridad y certeza a los usuarios acerca los posibles usos de la información contenida en los conjuntos de datos gubernamentales a ser

liberados y compartidos. Para efectos de la liberación de éstos y para facilitar su reutilización, deben respetarse al menos las siguientes condiciones básicas:

- Mantener el sentido original de la información.
- Siempre citar la fuente de la información.
- Explicitar la fecha de la última actualización de la información.

Consideración especial ha de tenerse presente en la liberación de datos gubernamentales, que contienen datos personales de ciudadanos, en el sentido de la protección de éstos. Del mismo modo han de resguardarse derechos de Propiedad Intelectual de terceros.

Se presentan los siguientes casos y las respectivas licencias asociadas a los contenidos:

1) La reutilización de obras protegidas por la propiedad intelectual se debe formalizar mediante el uso de la licencia de Reconocimiento Creative Commons “CC BY 3.0”. Este licenciamiento permite la reproducción, la distribución, la comunicación pública y la transformación para generar una obra derivada, sin ninguna restricción, siempre que se cite al autor.

Ver: <http://creativecommons.org/licenses/by/3.0/cl/>

2) Reconocimiento (by): Se permite cualquier explotación de la obra, incluyendo una finalidad comercial, así como la creación de obras derivadas, la distribución de las cuales también está permitida sin ninguna restricción.



3) En el caso de las fotografías, videos y otros tipos de datos en los cuales hay participación de terceros, la reutilización se debe realizar a través de la licencia Creative Commons Reconocimiento – SinObraDerivada CC BY-ND 3.0



La utilización, reproducción, modificación o distribución de los conjuntos de datos supone siempre la obligación de reconocer/citar/enlazar a la **[entidad publicadora]** como la fuente de los conjuntos de datos.

Ejemplo:

Fuente de los datos: Ministerio del Trabajo **[entidad publicadora]**

La aceptación de la licencia anteriormente descrita no supone la concesión de los derechos de autor ni la propiedad intelectual sobre los conjuntos de datos publicados por el organismo.

En determinados casos, hay que considerar que en Chile existen aún varios organismos que legalmente están autorizados a cobrar por el acceso a los datos que producen. Entre estos organismos se encuentran el SAF (Servicio Aerofotogramétrico de la Fuerza Aérea), Sernageomin (Servicio Nacional Geológico y Minero), Shoa (Servicio Hidrográfico y Oceanográfico de la Armada de Chile), IGM (Instituto Geográfico Militar), y Ciren (Centro de Información de Recursos Naturales de Chile).

6. Coordinación y Actualización

6.1 Hitos y Plazos

Hito	Plazo
Lanzamiento Portal data.gob.cl (Versión Beta)	Septiembre/2011
Período de Evaluación	Diciembre/2011 a Marzo/2012
Período de Regularización	Marzo/2012 a Septiembre/2012

Coordinación de Publicación y capacitaciones:

Área de Desarrollo (Unidad de Modernización y Gobierno Electrónico, MINSEGPRES).

6.2 Condiciones Especiales

Se presentan al menos dos condiciones especiales:

Caso 1. Instituciones que cuentan con conjuntos de datos de interés, de alta valoración, pero que no cuentan con los recursos (tecnológico o humanos) para ponerlos en línea en las condiciones

estipuladas en esta normativa. En estos casos la Unidad de Modernización y Gobierno Electrónico puede dar asistencia para suplir las dificultades mencionadas, dentro del Período de Regularización (6.1).

Caso 2. Instituciones que tienen datos publicados, pero que no están publicados de acuerdo a los estándares establecidos en la presente normativa. En este caso se tratará dentro del Plazo de Regularización (6.1).

6.3 Frecuencia de Actualización

A partir de la fecha de publicación de esta normativa, se establecerán dos períodos de actualización de la información contenida en el catastro, al final de cada semestre calendario. Se exceptuarán de esta medida aquellas instituciones cuya frecuencia de actualización de información sea anual.

Anexo 1 - Principios y Consideraciones

Los principios que sustentan la iniciativa Datos.gov.cl se basan en dos grandes ejes: las políticas y legislación sobre transparencia y privacidad; y los proyectos de publicación de datos gubernamentales en el mundo. Dichos principios, en cada uno de sus ejes, son:

a) Desde Transparencia y Privacidad:

- Relevancia
- Libertad de Información
- Apertura o transparencia
- Máxima divulgación
- Divisibilidad
- Facilitación
- No discriminación
- Oportunidad
- Principio del control
- Responsabilidad
- Gratuidad

En Chile, la Constitución Política⁴ consagra como regla general la publicidad de los actos y resoluciones de los órganos del Estado, así como los fundamentos y los procedimientos que utilicen (cuya excepción sólo puede emanar de una ley con quórum calificado que establezca el secreto o reserva cuando la publicidad de estos actos afecte el debido cumplimiento de las funciones de los órganos del Estado, los derechos de las personas, la seguridad de la nación o el interés nacional).

Adicionalmente y siguiendo la tendencia internacional, en abril de 2009 entra en vigencia la Ley de Transparencia N° 20.285 y su Reglamento, decreto N° 13 de 2009, que consagra la publicidad de los actos de los Órganos de la Administración del Estado y modifica las leyes del Congreso, de los tribunales, de la Contraloría General de la República, del Banco Central, del Ministerio Público, el Tribunal Constitucional y la Justicia Electoral, en el sentido que estas instituciones también queden regidas por el principio de la transparencia. 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, a través de los medios y procedimientos que al efecto establezca la ley.

Por otra parte, la principal regulación en materia de privacidad en Chile está contenida en la ley N° 19.628 sobre Protección a la Vida Privada, la cual regula el tratamiento de los datos personales y define conceptos fundamentales como datos de carácter personal y datos

⁴ Ver http://www.bcn.cl/lc/cpolitica/index_html

sensibles, junto con establecer la oportunidad del tratamiento de datos personales, los derechos de los titulares, el tratamiento por parte de los organismos públicos, así como las sanciones por infracciones a la ley.

b) Desde las políticas de apertura de datos internacionales

- Énfasis en el acceso
- Plataforma abierta
- Datos desagregados (raw data)
- Crecer y mejorar utilizando el feedback (comunidades)
- Responsabilidad distribuida (velar por secretos, calidad, contexto, entre otros)
- Incorporación rápida
- Buenas prácticas
- Localizado y enfatizado en norma
- Alineado con privacidad y legislación

La publicación de datos en forma abierta instituye un nuevo modelo de negocios sobre los datos. Gracias a las posibilidades tecnológicas ofrecidas por la red Internet, por la Web como capa lógica sobre esa red, y por las capacidades de hardware y software existentes, se ha generalizado a nivel internacional el movimiento de publicación de datos abiertos, es decir, de exponer pública y abiertamente conjuntos de datos en formatos accesibles, para ser utilizados de manera simple y fácil por cualquier usuario para los fines que estime conveniente.

Este movimiento tecnológico y social ha motivado la creación de sistemas de publicación abierta por parte de los gobiernos en diferentes países. También ha impulsado el desarrollo de tecnologías y protocolos basado en diversos estándares internacionales, y apoyadas por agencias que los proveen (por ejemplo, ISO, W3C, OASIS, NIST, IETF, etc.)

Los beneficios que a su vez se esperan de una iniciativa como Datos.gob.cl, están en directa relación con los principios antes mencionados y en líneas generales, se pueden mencionar:

- Mejorar políticas de transparencia
- Mejorar la participación ciudadana
- Agregar valor a la información del Estado
- Mejorar la calidad de los datos a través del escrutinio público
- Promover la innovación
- Promover el crecimiento económico
- Mejorar la eficiencia de la gestión pública y calidad de sus políticas

Anexo 2 - Referencias Normativas

Leyes y Normativas pre-existentes que dan contexto a esta iniciativa de datos abiertos en el caso de Chile.

- Ley N° 20.285 de 2008 del Ministerio Secretaría General de la Presidencia sobre acceso a la información pública, y su Reglamento contenido en el Decreto N° 13 de 2009. Regula el principio de transparencia de la función pública, el derecho de acceso a la información de los órganos de la Administración del Estado, los procedimientos para el ejercicio del derecho y para su amparo, y las excepciones a la publicidad de la información.
- Ley N° 17.374 de 1970 del Ministerio de Hacienda, que crea el Instituto Nacional de Estadísticas (INE), y su Reglamento, contenido en el decreto N° 1.062 de 1970. Regula la naturaleza y atribuciones del INE y de sus principales organismos, así como el Secreto Estadístico, el cual establece la prohibición del INE y de los organismos fiscales, semifiscales y Empresas del Estado, y cada uno de sus funcionarios, de divulgar los hechos que se refieren a personas o entidades determinadas de que hayan tomado conocimiento en el desempeño de sus actividades. Los datos estadísticos no podrán ser publicados o difundidos con referencia expresa a las personas o entidades a quienes directa o indirectamente se refieran, si mediare prohibición del o los afectados.
- Ley N° 19.628 de 1999 del Ministerio Secretaría General de la Presidencia sobre protección a la vida privada. Regula el tratamiento de los datos personales y define conceptos fundamentales como datos de carácter personal y datos sensibles, junto con establecer la oportunidad del tratamiento de datos personales, los derechos de los titulares, el tratamiento por parte de los organismos públicos, así como las sanciones por infracciones a la ley.
- Decreto Supremo N° 77 de 2004 del Ministerio Secretaría General de la Presidencia, aprueba norma técnica sobre eficiencia de las comunicaciones electrónicas entre órganos de la Administración del Estado y entre éstos y los ciudadanos. Regula de manera general y supletoria las comunicaciones por medios electrónicos entre los órganos de la Administración del Estado y entre éstos y los ciudadanos, debiendo someterse a los principios de legalidad, efectividad, eficiencia, publicidad y transparencia.
- Decreto Supremo N° 81 de 2004 del Ministerio Secretaría General de la Presidencia, aprueba norma técnica para los órganos de la Administración del Estado sobre interoperabilidad de documentos electrónicos. Regula las características mínimas obligatorias de interoperabilidad que debe cumplir los documentos electrónicos en su generación, envío, recepción, procesamiento y almacenamiento; y es aplicable a los documentos electrónicos que se generen, intercambien, transporten y almacenen en o entre los diferentes órganos de la Administración del Estado y en las relaciones de éstos con los ciudadanos, cuando tengan lugar utilizando técnicas y medios electrónicos.

- Decreto Supremo N° 83 de 2004 del Ministerio Secretaría General de la Presidencia, aprueba norma técnica para los órganos de la Administración del Estado sobre seguridad y confidencialidad de los documentos electrónicos. Regula las características mínimas obligatorias de seguridad y confidencialidad, así como las recomendaciones. Las exigencias y recomendaciones previstas en esta norma, tienen por finalidad garantizar estándares mínimos de seguridad en el uso, almacenamiento, acceso y distribución del documento electrónico; facilitar la relación electrónica entre los órganos de la Administración del Estado y entre éstos y la ciudadanía y el sector privado en general; y salvaguardar el uso del documento electrónico de manera segura, confiable y en pleno respeto a la normativa vigente sobre confidencialidad de la información intercambiada.
- Decreto Supremo N° 100 de 2004 del Ministerio Secretaría General de la Presidencia, aprueba norma técnica para el desarrollo de sitios web de los órganos de la Administración del Estado. Regula las características mínimas obligatorias que se deben cumplir y aquellas que se recomiendan en los sitios web de los órganos de la Administración del Estado. Éstos deben ser desarrollados de tal forma que garanticen la disponibilidad y accesibilidad de la información, así como el resguardo de los derechos de los titulares de datos personales y asegurando la interoperabilidad de los contenidos, funciones y prestaciones ofrecidas por el órgano de la Administración, con prescindencia de las plataformas, hardware y software utilizados.

Anexo 3 - Términos y Definiciones

Base de Datos Conjunto de datos con significados inherentes construido con algún propósito específico y dirigida a un grupo de usuarios específicos.

Catálogo (de datasets): Relación ordenada de términos y conceptos que describen los metadatos de un conjunto de datos. Usado generalmente para ordenar y localizar la información por parte de los usuarios.

Conjunto de datos: Datos relacionados, convenientemente estructurados y organizados, de forma que puedan ser tratados (procesados) apropiadamente para obtener información.

Conjunto de metadatos estructurados asociados que describe el significado de los datos.

Consumidor (de datos): Cualquier persona u organización que accesa datos publicados y obtiene una copia total o parcial de ellos para fines propios.

Dataset (Conjunto de datos): Colección de datos estructurados con unidad temática y física. No está necesariamente dirigido a un usuario específico.

Dato: Representación simbólica (numérica, alfabética, algorítmica etc.), atributo o característica de una entidad. Un dato es una expresión mínima de contenido sobre un tema.

Datos Abiertos: (*Open data*) Datos que no tienen restricciones de acceso de ningún tipo, en particular ni administrativas ni tecnológicas.

Datos Enlazados: (*Linked data*) Objetos de información que están enlazados por protocolos informáticos, en particular por medio de URIs o enlaces href (Web).

Datos personales: toda información numérica, alfabética, gráfica, fotográfica, acústica o de cualquier otro tipo sobre una persona física identificada o identificable, cuya identidad pueda determinarse, directa o indirectamente mediante un número de identificación o uno o varios elementos específicos, característicos de su identidad física, fisiológica, psíquica, económica, cultural o social.

Derechos de autor: Derecho moral, irrenunciable e inalienable, y derecho exclusivo de explotación que tiene y ejerce sobre su obra literaria, artística o científica la persona creadora de la misma.

Derechos de propiedad intelectual: derechos de autor y derechos afines, incluidas las formas de protección sui géneris.

Diccionario (de datos): Descripción del esquema y taxonomías usadas en los conjuntos de datos.

Disponibilidad: Propiedad o característica de los activos consistente en que las entidades o procesos autorizados tienen acceso a los mismos cuando lo requieren.

Documento: Toda información cualquiera que sea su soporte material o electrónico así como su forma de expresión gráfica, sonora o en imagen utilizada.

Esquema (de datos): definición de conjunto de campos de diferentes tipos que tienen por objetivo registrar y organizar metadatos

Formato (de datos): Estructura física y lógica usada para almacenar datos en un archivo. Usualmente se indica por un sufijo al final del nombre del archivo. Ejemplo: miArticulo.pdf, publicaciones.xml.

Formato: Conjunto de características técnicas y de presentación.

Infomediario/a: empresa o negocio cuyo modelo se base en administrar información en favor de terceros, por medio de la recopilación de datos de diversas fuentes, su estudio y selección, para organizarlos y distribuirlos de manera relevante para su clientela.

Información: conocimientos comunicados o adquiridos que permiten ampliar o precisar los que se poseen sobre una determinada materia.

Información: Hecho conocido, con significado explícito. La información consta de datos más metadatos que los describen.

Metadatos: Datos y/o documentos que describen otros datos en términos de contexto, contenido o, en general, cualquier aspecto que se considere necesario para ampliar la conceptualización del dato descrito.

Publicación de Datos: Modelo de intercambio de datos abierto, que no requiere acuerdos bilaterales previos entre el publicador del dato y el consumidor del dato. Los datos están disponibles bajo formatos y patrones de proceso conocidos y estándares. El objetivo es que cualquier contraparte pueda usar los datos y metadatos.

Publicador: La agencia o individuo que se hace responsable por la publicación de datos.

Raw data: Expresión que refiere a datos en estado “original”, no derivados de otro conjunto de datos, es decir, previo a su procesamiento o agregación.

Repositorio (de datos): Unidad virtual de los conjuntos de datos físicos publicados por una organización.

Servicio público: actividad llevada a cabo por la Administración o, bajo un cierto control y regulación de esta, por una organización, especializada o no, y destinada a satisfacer necesidades de la colectividad.

Sistema: Conjunto de soluciones tecnológicas (informáticas) que permiten la organización, almacenamiento y recuperación de datos.

Transparente: Claro, evidente, que se comprende sin duda ni ambigüedad.

URI (Uniform Resource Identifier): En español “identificador uniforme de recurso”, es una cadena de caracteres corta que identifica inequívocamente un recurso (servicio, página, documento, dirección de correo electrónico, enciclopedia, etc.). Normalmente estos recursos son accesibles en una red o sistema. Los URI pueden ser localizadores uniformes de recursos (URL), Uniform Resource Name (URN), o ambos (URL + URN).

URL: (Uniform Resource Locator). Nombre/Identificador. Secuencia compacta de caracteres que permite localizar un recurso describiendo su modo primario de acceso. Las URL son un subconjunto de las URI.

Anexo 4 - Abreviaturas y Acrónimos

Las URL referencian información complementaria y/o especificaciones.

API (Application Programming Interface): es una interfaz de comunicación entre componentes de software, que ofrece un conjunto de llamadas a ciertas librerías de programación que dan acceso a ciertos servicios desde los procesos, consiguiendo la abstracción en la programación entre niveles inferiores y superiores del software.

CSV: Comma separated Values. <http://www.ietf.org/rfc/rfc4180.txt>

DCAT (Data Catalog Vocabulary) : Vocabulario RDF para la interoperabilidad de catálogos de datos. Su objetivo principal es la expresión de los catálogos de datos de gobiernos en un formato estándar usando RDF.

DCMI: Dublin Core Metadata Initiative. <http://dublincore.org/documents/dces/>

GML: Geography Markup Language. <http://opengis.net/gml/>

HTML: Hyper Text Markup Language. <http://www.w3.org/html/>;
http://www.w3.org/egov/wiki/Data_Catalog_Vocabulary

JPEG: Join Photographic Expert Group. <http://www.jpeg.org/>

KML: Keyhole Markup Language. <http://www.opengeospatial.org/standards/kml/>

MIME (Multipurpose Internet Mail Extensions): una serie de convenciones o especificaciones dirigidas al intercambio a través de Internet de todo tipo de archivos (texto, audio, vídeo, etc.) de forma transparente para el usuario.

MPEG-7: Moving Picture Expert Group. <http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm>

PNG: Portable Network Graphics. <http://www.libpng.org/pub/png/>

RDF (Resource Description Framework) : Infraestructura para describir semánticamente recursos, es decir, dotar de sentido a lo que representamos para que las máquinas lo comprendan. RDF se puede representar en distintos formatos: XML, N3, Turtle, etc. www.w3.org/RDF

RDFa (Resource Description Framework-in-attributes): Forma de representar los datos estructurados visibles en las páginas Web mediante unas anotaciones semánticas, incluidas en el código e invisibles para el usuario, que permiten a las aplicaciones interpretar esta información y utilizarla de forma eficaz. <http://www.w3.org/TR/2008/WD-xhtml-rdfa-primer-20080620/>

REST (Representational State Transfer): cualquier interfaz web simple que utiliza XML y HTTP, sin las abstracciones adicionales de los protocolos basados en patrones de intercambio de mensajes como el protocolo de servicios web SOAP.

Reutilización pública: uso de documentos que obran en poder del sector público, por personas físicas o jurídicas, con fines comerciales o no comerciales, siempre que dicho uso no constituya una actividad administrativa pública.

Reutilizar: utilizar algo, bien con la función que desempeñaba anteriormente o con otros fines.

SDMX: Statistical Data and Metadata eXchange. www.sdmx.org

SKOS (Simple Knowledge Organization System): vocabulario RDF para la representación de sistemas de conocimiento semi-formales, tales como tesauros, taxonomías y esquemas de clasificación. SKOS se ha diseñado para facilitar la migración de los sistemas organizacionales existentes a la Web Semántica.

SPARQL (SPARQL Protocol and RDF Query Language): tecnología de consulta de información desde bases de datos y otros orígenes de datos en sus estados primitivos a través de la Web. Se compone de un lenguaje de consulta estandarizado y de un protocolo con el que ofrecer un servicio Web estándar (HTTP / SOAP), permitiendo realizar consultas a diversas fuentes de datos que almacenan los mismos nativamente en RDF o los presentan como tal.

SPSS: Statistical Package for Social Science. <http://www.spss.com/>

SSL/TLS: Secure Socket Layer/Transport Layer Security.
<http://datatracker.ietf.org/wg/tls/charter/>

STATA: Data Analysis and Statistical Software. www.stata.com

SVG: Scalable Vector Graphics. <http://www.w3.org/Graphics/SVG/>

TAR: Tape Archiv.
<http://www.freebsd.org/cgi/man.cgi?query=tar&sektion=5&manpath=FreeBSD+8-current>

URI: Uniform Resource Identifier. <http://tools.ietf.org/html/rfc3986>

W3C (World Wide Web Consortium): es un consorcio internacional de reconocido prestigio donde las organizaciones miembro, el personal a tiempo completo y el público en general, trabajan conjuntamente para desarrollar estándares Web.

W3C World Wide Web Consortium. www.w3.org

WGS84: World Geodetic System 1984. http://earth-info.nga.mil/GandG/publications/tr8350.2/tr8350_2.html

XHTML: eXtensible Hyper Text Markup Language. <http://www.w3.org/TR/xhtml1/>

XML: eXtensible Markup Language. <http://www.w3.org/TR/2000/REC-xml-20001006>

Anexo 5 - Referencias

- OPEN GOVERNMENT: Gobierno Abierto, César Calderón Sebastián Lorenzo, ALGÓN EDITORES 2010
- CTIC, Glosario Open Data, <http://datos.fundacionctic.org/glosario-open-data/>
- W3C, Publishing Open Government Data, <http://www.w3.org/TR/gov-data/>
- Open Data Navarra, http://www.navarra.es/home_es/Open-Data/
- Open Data Euskadi, <http://opendata.euskadi.net/w79-home/es>