

BORRADOR NORMATIVA TÉCNICA SOBRE PUBLICACIÓN DE DATOS ABIERTOS

CAPÍTULO I

Disposiciones Generales

Artículo 1°.- La norma técnica sobre publicación de datos abiertos establecidas en el presente decreto serán aplicables a los órganos de la Administración del Estado que se encuentran señalados en el inciso segundo del artículo 1° del decreto con fuerza de ley N°1, de 2001, del Ministerio Secretaría General de la Presidencia, que fija texto refundido, coordinado y sistematizado de la ley N° 18.575, Orgánica Constitucional de Bases Generales de la Administración del Estado, con excepción de las empresas públicas creadas por ley.

Artículo 2°.- Para los efectos de la norma técnica establecida en este decreto, se entenderá por:

Base de datos: Conjunto de datos organizados, recopilados y contruidos con algún fin determinado.

Catálogo de conjunto de datos o 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 o Dataset: datos relacionados, estructurados y organizados de forma que puedan ser tratados o procesados apropiadamente para obtener determinada información. Incorporan un conjunto de metadatos asociados que describe el significado de los datos.

Consumidor de datos: cualquier persona u organización que acceda a datos publicados y obtiene una copia total o parcial de ellos para fines propios.

Dato: representación simbólica de tipo numérica, alfabética y algorítmica, entre otros, atributo o característica de una entidad. Un dato es una expresión mínima de contenido sobre un tema.

Dato abierto: dato que no tienen restricciones de acceso de ningún tipo, en particular ni administrativas, ni legales, ni tecnológicas, permitiendo su reutilización.

Datos estadísticos: datos usados por usuarios especializados del área estadística. Se encuentran en formatos particulares para su reutilización y explotación de aplicaciones orientadas al trabajo estadístico.

Datos enlazados: objetos de información que están enlazados por protocolos informáticos, en particular por medio de diversos URI o enlaces (links).

Datos genéricos: datos respecto de los que se carece de aplicaciones especializadas.

Datos geográficos: datos que provienen de, o son de uso en, aplicaciones geográficas. Generalmente se encuentran almacenados en formatos específicos para representar la superficie terrestre o la geometría de objetos en el espacio geográfico.

Datos históricos: datos que corresponden a una fecha anterior a la de actual publicación.

Dato personal: Datos de carácter personal o datos personales, relativos a cualquier información concerniente a personas naturales, identificadas o identificables.

Dato sensible: aquellos datos personales que se refieren a las características físicas o morales de las personas o a hechos o circunstancias de su vida privada o intimidad, tales como los hábitos personales, el origen racial, las ideologías y opiniones políticas, las creencias o convicciones religiosas, los estados de salud físicos o psíquicos y la vida sexual.

Disponibilidad: propiedad consistente en que las entidades o procesos autorizados tienen acceso a los datos cuando lo requieran.

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

Escala: relación matemática que existe entre las dimensiones reales y las del dibujo que representa la realidad sobre un plano o un mapa.

Formato: conjunto de características técnicas y de presentación.

Formato abierto: formato que permite la comprensión y reutilización de datos tanto por personas como por sistemas automatizados.

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.

Imagen: figura o representación visual de la apariencia de una persona u objeto.

Información: hecho conocido, que posee por características un significado explícito.

Metadatos: datos 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.

Protocolos de pre-procesamiento: procedimiento de manejo de información que cambian la naturaleza inicial del dato y su contenido.

Publicación de datos: modelo de intercambio de datos abiertos, 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.

Publicador: la entidad que se hace responsable por la publicación de datos.

Repositorio de datos: unidad virtual de los conjuntos de datos publicados por uno o varios publicadores.

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

UTF-8 (*UCS Transformation Format 8*): formato de codificación de caracteres Unicode e ISO 10646 utilizando símbolos de longitud variable, RFC 3629.

URI (*Uniform Resource Identifier*): identificador uniforme de recurso, es una cadena de caracteres corta que identifica inequívocamente un recurso (servicio, página, documento, dirección de correo electrónico, enciclopedia, etc.), RFC 3986.

URL (*Uniform Resource Locator*): Nombre o identificador que refiere a una secuencia compacta de caracteres que permite localizar un recurso describiendo su modo primario de acceso. Las URL son un subconjunto de las URI.

CAPÍTULO II

Características de los datos abiertos.

Artículo 3º.- Los datos para ser considerados como abiertos, deberán contar como mínimo con las siguientes características:

i. Completos: Deben proporcionarse todos los datos con la única excepción de aquellos que cuenten con especial protección de la vida privada, seguridad o que por ley se señale expresamente.

- ii. Primarios: Los datos deben compartirse tal y como se originan y con el mismo nivel de detalle, sin modificaciones, agregados u otras operaciones intermedias adicionales.
- iii. Oportunos: Los datos deben proporcionarse prontamente, sin retrasos en su publicación, para que no pierdan su valor por quedar obsoletos.
- iv. Accesibles: Los datos deben estar disponibles de forma igualitaria y equitativa para todos los usuarios.
- v. Legibles por máquinas: Los datos deben estar estructurados de forma que permitan su tratamiento automatizado.
- vi. No discriminatorios: El acceso a los datos debe ser abierto e inmediato sin requisitos adicionales de registro o identificación.
- vii. Formatos libres: Los datos se publican en formatos no propietarios para maximizar su potencial de reutilización.
- viii. En formatos abiertos: Los datos se comparten libres de derechos, de forma que cualquiera pueda usarlos, reutilizarlos y redistribuirlos sin restricciones.

CAPÍTULO III

Condiciones previas

Artículo 4°.- Los órganos de la Administración del Estado deberán publicar sus datos abiertos o datasets en el repositorio central de datos públicos administrado por el Ministerio Secretaría General de la Presidencia cuya URL es <http://datos.gob.cl>, denominado “Repositorio Central”, asegurando las condiciones operacionales para garantizar la continuidad de los servicios. Sin perjuicio de lo anterior, cada institución podrá contar con su propio repositorio de datos abiertos, debiendo ser replicado en el repositorio central de datos públicos. Se entenderá por replicado que todas las publicaciones en el portal institucional deben ser publicadas al mismo tiempo en el Repositorio Central, para garantizar una única versión de los datos.

Se deja constancia que cada institución será responsable y tendrá el control de los datos o datasets que publiquen en el repositorio de datos públicos.

Artículo 5°.- Cada institución, a través de su jefe de servicio, deberá designar el área, equipo o persona encargada, dentro de los 30 días hábiles contados desde la entrada en vigencia del presente Decreto, de la publicación de datos abiertos de su institución en el Repositorio Central y de asegurar la integridad, autenticidad, sistema de licenciamiento y permanencia en el tiempo de los datos publicados por la institución.

Artículo 6°.- Antes de la publicación de los datos la institución deberá procurar que éstos cumplan con las siguientes condiciones generales:

1. No publicar datos personales o datos sensibles según lo establecido en la Ley N° 19.628 sobre Protección de la Vida Privada u otra que a futuro la reemplace, salvo las excepciones establecidas en la mencionada ley.
2. Procurar que los datos a publicar mantengan su calidad y fiabilidad, lo que quiere decir que los datos a publicar tengan un contenido razonablemente confiable y adecuado para ser directamente reutilizado.
3. Los datos deberán ser publicados previos a cualquier agregación estadística o pre-procesamiento, salvo cuando se trate de datasets que contengan datos personales o datos sensibles, los que deberán procesarse y disociarse previo a su publicación, en cumplimiento de la ley N° 19.628.
4. Los datos deben estar en formatos abiertos, no propietarios, estables y ser accesibles por diversos medios o plataformas.
5. Considerar su disponibilidad en múltiples formatos.
6. Cada dataset debe tener un identificador único y permanente, según el estándar URI.
7. 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.
8. Los metadatos obligatorios a utilizar son los definidos en esta norma.
9. En los casos de que los dataset estén constituidos por uno o más recursos, cada dataset completo debe ser posible de reconstruir a partir de los metadatos que se describen en la distribución del dataset.
10. Para dataset con validez temporal, se debe indicar desde cuándo y hasta cuándo son válidos.
11. Para dataset con actualización periódica, se debe indicar la periodicidad de actualización.
12. Los metadatos para coberturas y datos georreferenciados deben contener los campos de metadatos obligatorios recomendados por el Sistema Nacional de Información Territorial del Ministerio de Bienes Nacionales (SNIT) o por lo que se rija la institución.
13. Para datos estadísticos se debe incluir los metadatos correspondientes a metodología y calidad de datos.
14. Para asegurar la compatibilidad en la codificación de caracteres en sistemas digitales se utilizará la codificación UTF-8.
15. Velar por el secreto estadístico de los datos publicados, según lo que señala la ley N° 17.374, que fija nuevo texto refundido, coordinado y actualizado del decreto con fuerza de ley N° 313 de 1960, que aprobara la ley orgánica dirección estadística y censos y crea el Instituto Nacional de Estadísticas.
16. Velar por el cumplimiento de otras leyes sectoriales o especiales.

CAPÍTULO IV

Formatos de publicación

Artículo 7°.- Datos genéricos.

Para los datos genéricos se debe usar formatos estándares y abiertos. En caso que se trate de datos históricos o la transformación a formatos abiertos sea excesivamente oneroso para la institución, podrán usar formatos propietarios o aún no estandarizados, indicando el software, la versión y productor que permite procesar esos formatos.

Artículo 8°.- Datos geográficos.

Se recomienda que los formatos de publicación de coberturas y datos georreferenciados estén en concordancia con los formatos definidos por el SNIT, sin perjuicio de las instituciones que se rijan por otros estándares.

En caso en que más de una institución ejerza potestades sobre un mismo territorio, éstas deberán coordinarse de manera interna, para utilizar formatos y escalas que sean iguales o comparables.

Artículo 9°.- Imágenes y datos estadísticos.

Para la publicación de imágenes y datos estadísticos se deben utilizar formatos y licencias que permitan su reutilización, prefiriendo aquellos de uso común a nivel internacional.

CAPÍTULO V

Proceso de Publicación

Artículo 10.- Etapas del proceso de publicación.

Se deben considerar las siguientes etapas o fases:

1. Levantamiento de las bases de datos que posea la institución.
2. Definición y decisión sobre publicación de dataset.
3. Preparación de los datos y metadatos del dataset.
4. Verificación y control de las características de la publicación.
5. Publicación de datos y aseguramiento de disponibilidad del dataset.

Artículo 11.- Lenguaje.

El lenguaje que se debe utilizar para los datos y metadatos es Español-Chile (ES_CL) y la codificación de caracteres debe ser UTF-8.

Artículo 12.- Metadatos.

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

1. **Título:** nombre definido para dataset.

2. **Descripción:** resumen descriptivo del dataset, especificando el contenido de recurso.
3. **Institución:** nombre de la institución o entidad responsable de hacer el dataset disponible al público y el correo electrónico del área, equipo o persona encargada de su publicación.
4. **Licencia:** información sobre los derechos de uso del dataset.
5. **Categoría:** clasificación conceptual básica del dataset en sistemas de categorías o taxonomías disponibles.
6. **Fecha de publicación:** punto de tiempo asociado al evento de publicación del dataset. Se asigna de manera automática.
7. **Periodicidad de la actualización:** frecuencia con la cual el dataset es actualizado.
8. **Recurso:** contenido y características físicas de un dataset. Un mismo dataset puede tener varios recursos, siempre y cuando se busque publicar la misma información utilizando distintos formatos.

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

- a) **Enlace de distribución:** URL del dataset.
- b) **Tamaño:** para el caso de archivos, volumen físico utilizado, en unidades estándares. Se asigna de manera automática.
- c) **Formatos:** modos o medios principales de representación de los datos, normalmente identificados en la extensión del nombre del archivo. Se asigna de manera automática.
- d) **Descripción:** características adicionales tales como tipo de compresión, versión de algún estándar, entre otras, necesarias de conocer para su procesamiento.
- e) **Fecha de los metadatos:** fecha de publicación.

Artículo 13.- Licencias.

La información que sea publicada en el repositorio de datos públicos debe ser de libre uso, esto es, que no tenga barreras para su reutilización y difusión. Para ello se deben utilizar licencias abiertas de modo que se permita el procesamiento, el almacenamiento y la republicación de la información para cualquier tipo de uso.

CAPÍTULO VI

Catálogo

Artículo 14.- Existirá un catálogo central que será administrado por el Ministerio Secretaría General de la Presidencia, sin perjuicio de lo anterior, cada órgano de la administración del Estado podrá tener un catálogo institucional. En caso que las instituciones públicas cuenten con un catálogo propio éste deberá ser concordante con el catálogo central, esto es, tener entradas en el catálogo central para cada uno de los dataset que posea.

Artículo 15.- En el caso de contar con un catálogo institucional deberá cumplir con las siguientes condiciones:

1. Contener un listado exhaustivo de los dataset presentes en el repositorio.

2. Permitir la navegación, búsqueda y consulta simple de los datos contenidos en el repositorio. Para esto implementará todas las funcionalidades necesarias que se requieran, tales como categorizaciones, navegación, búsqueda de texto, lenguaje de consulta.
3. Los catálogos utilizarán un conjunto de metadatos mínimos definidos en esta norma y los extenderán con estándares de uso internacional o nacional.
4. El catálogo debe proveer funcionalidades de retroalimentación de los usuarios, así como consulta y sugerencias.
5. Debe estar enlazado con el catálogo central.
6. El responsable del catálogo debe velar por su funcionamiento, mantener su disponibilidad y analizar sus estadísticas de uso.

Artículo 16.- Los jefes de servicio serán responsables de velar por el cumplimiento de lo dispuesto en la presente norma técnica.

CAPÍTULO VII

Disposiciones finales

Artículo 17.- El Ministerio Secretaría General de la Presidencia, mediante resolución, podrá dictar guías técnicas para facilitar la comprensión e implementación de las normas técnicas a la que se refiere el presente decreto. Las guías técnicas que se dicten al efecto, deberán ser puestas a disposición del público en el sitio web www.guiadigital.gob.cl.

Artículo 18.- Según lo estipulado en el artículo 47 del Decreto Supremo N° 181, de 2002, del Ministerio de Economía, Fomento y Turismo, el Ministerio Secretaría General de la Presidencia deberá publicar en su sitio web, semestralmente, el nivel de cumplimiento de las normas fijadas en virtud del presente decreto.

Disposiciones transitorias

Artículo 1°.- Las disposiciones de este Decreto, comenzará a regir un año después de su publicación en el Diario Oficial. Sin perjuicio de lo anterior, los órganos de la Administración del Estado, dentro del plazo de seis meses desde la publicación en el Diario Oficial del presente Decreto, deberán confeccionar un plan de publicación que contendrá un listado de bases de datos a publicar con las siguientes categorías:

1. Título de las bases de datos;
2. El área específica del organismo que genera dicha base de datos;
3. La periodicidad de la actualización de los datos;
4. Fecha aproximada, con indicación de mes y año, de publicación en el repositorio central de datos públicos.