
 IDERA Infraestructura de Datos Espaciales de la República Argentina	Tipo de documento: Documento teórico	Versión: 1.0
	Grupo de trabajo: Metadatos	Fecha : 15/11/2017
	Guía de Buenas Prácticas de Metadatos	Página 1 de 17




IDERA
Infraestructura de
Datos Espaciales de la
República Argentina

Guía de Buenas Prácticas de Metadatos


	Tipo de documento: Documento teórico	Versión: 1.0
	Grupo de trabajo: Metadatos	Fecha : 15/11/2017
	Guía de Buenas Prácticas de Metadatos	Página 2 de 17

<u>Título</u>	Guía de Buenas Prácticas de Metadatos
<u>Versión</u>	1.0
<u>Autor</u>	Metadatos
<u>Fecha</u>	15/11/2017
<u>Estado</u>	En revisión
<u>Publicador</u>	Grupo de Trabajo de Metadatos de IDERA
<u>Contacto</u>	coordinador_metadatos@idera.gob.ar
<u>Resumen</u>	Este documento contiene la descripción de referencia al uso de buenas prácticas para la creación, análisis y procesamiento de metadatos, basándonos en estándares como Normas ISO y OGC.
<u>Palabras claves</u>	Buenas prácticas, recomendaciones, fuentes de información, metadatos, IDERA.

	Tipo de documento: Documento teórico	Versión: 1.0
	Grupo de trabajo: Metadatos	Fecha : 15/11/2017
	Guía de Buenas Prácticas de Metadatos	Página 3 de 17

Índice

Registro de Versión	4
Objetivo	4
Alcance	4
Introducción	4
Desarrollo	5
Proceso de generación de Metadatos para un producto	5
Paso 1: Documentar Preguntas	6
Paso 2: Analizar contenido	6
Paso 3: Establecer estándar adicionales	7
Paso 4: Generar XML	7
Paso 5: Gestionar la publicación	8
Consistencia de geoservicios para evitar la duplicidad de información espacial y metadatos	8
Implementación de jerarquía y agrupamiento de fuentes de datos, únicas y múltiples	9
Tratamiento de productos derivados	10
Formato de intercambio de metadatos	11
Restricciones de uso y licencias	11
Recomendaciones Generales	11
Referencias bibliográficas	16

	Tipo de documento: Documento teórico	Versión: 1.0
	Grupo de trabajo: Metadatos	Fecha : 15/11/2017
	Guía de Buenas Prácticas de Metadatos	Página 4 de 17

Registro de Versión

A continuación detallaremos el versionado del documento:

Fecha	Versión	Descripción
15/11/2017	1.0	Creación del documento.-

Objetivo


Lograr una implementación estandarizada del uso de buenas prácticas para la creación, análisis, producción, carga y distribución de metadatos tanto para la información primaria, como de productos derivados.

Alcance

El presente documento es un estándar implementado para la descripción, documentación y catalogación de recursos de información utilizada en el proyecto IDERA, y recomendado para su utilización en otros ámbitos públicos y privados de la Argentina.

Introducción

Durante el tratamiento de este documento se han incorporado detalles conceptuales y operativos con la intención de estandarizar procesos comunes tal que permitan una simple y generalizada comprensión, y homogeneización de la implementación de las buenas prácticas con respecto a la producción, análisis, manejo, intercambio y publicación de metadatos.

	Tipo de documento: Documento teórico	Versión: 1.0
	Grupo de trabajo: Metadatos	Fecha : 15/11/2017
	Guía de Buenas Prácticas de Metadatos	Página 5 de 17

Desarrollo

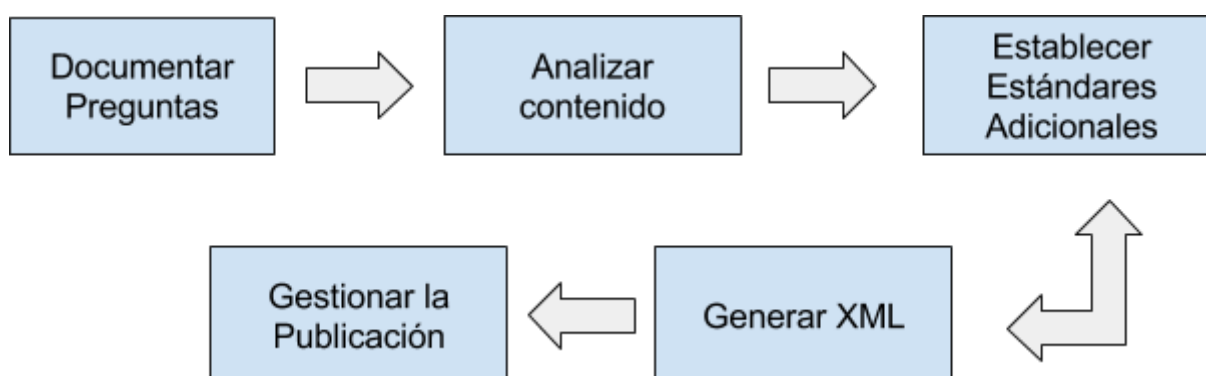
A continuación se describen una serie de conceptos a tener en cuenta en el proceso de generación y manipulación de metadatos.


Proceso de generación de Metadatos para un producto

El proceso de documentación de los conjuntos de datos científicos ha cambiado a lo largo de las últimas décadas con la aparición de los estándares de metadatos, se está tratando de explicar las necesidades de documentación de los mismos para poder lograr unicidad en la información generada, muchas veces costosa que posteriormente incrementa el costo al momento de generar búsquedas y actualización de los datos.

Pero surge en este proceso la dificultad de “encajar” los datos científicos en estándares como el adoptado, lo que nos lleva a realizar un grupo de pasos para adecuar estos datos y así llegar a un resultado exitoso.

Por la experiencia adquirida se ha aprendido que el enfoque a aplicar tiene que ser según la lógica ABM (alta, baja y modificación) de metadatos que como se ilustra en la siguiente figura debe recorrer los siguientes pasos:



	Tipo de documento: Documento teórico	Versión: 1.0
	Grupo de trabajo: Metadatos	Fecha : 15/11/2017
	Guía de Buenas Prácticas de Metadatos	Página 6 de 17


Paso 1: Documentar Preguntas

El proceso se inicia con las necesidades de documentación de un producto, teniendo en cuenta las siguientes preguntas sencillas, como por ejemplo:

- Posee diferentes categorías de datos?
- Existen diferentes subgrupos temporales y espaciales?
- Tiene conjuntos de datos con múltiples fuentes?
- Es necesario hacer referencia a Recursos en línea?
- Qué se necesita para describir una serie de gránulos relacionados?
- Es necesario describir tipos de agregaciones?
- Varía la calidad de datos en el conjunto de datos?
- Qué se necesita para realizar un seguimiento de procesamiento de múltiples fuentes de datos?
- Qué se necesita para realizar un seguimiento de cumplimiento de las normas?
- Es necesario utilizar las características espaciales para describir la calidad, como los indicadores de la calidad?
- Es necesario explicar por qué se realizaron procedimientos a los datos?
- Tiene conjuntos de datos en múltiples ubicaciones?
- Es necesario para describir los instrumentos utilizados para hacer observaciones?
- Es necesario identificar inequívocamente los espacio de nombres utilizados?
- Quieres servir a los metadatos utilizando un servicio web REST(Representational State Transfer)?
- Es necesario hacer un seguimiento de los problemas de usuario?
- Qué se necesita para realizar un seguimiento de los requisitos y planes?
- Qué se necesita para realizar un seguimiento de las transformaciones y de procesamiento de datos?

Paso 2: Analizar contenido

Una vez que las preguntas son conocidas para un determinado conjunto de datos y la información necesaria para responder a las preguntas puede ser determinada, tenemos que

	Tipo de documento: Documento teórico	Versión: 1.0
	Grupo de trabajo: Metadatos	Fecha : 15/11/2017
	Guía de Buenas Prácticas de Metadatos	Página 7 de 17

analizar cuáles de los elementos se adapta dentro del modelo UML de la ISO para poder representar de manera lógica y unificada el perfil de metadatos a utilizar para este producto específico.


Paso 3: Establecer estándar adicionales

Una vez que se conocen los componentes de información, podemos considerar la cuestión de cómo la información de estos componentes podría representarse mediante diversas normas, en nuestro caso como aplicamos ISO 19115, muchas veces podemos encontrarnos que los elementos utilizados no nos permiten estructurar la información. Por esto tenemos que acudir a otro tipo de estándares para casos como la determinación de Tesoros al momento de generar palabras claves, el uso de protocolos de servicios estandarizados como los determinados por OGC (Open Geospacial Consortium) para la descripción de un instrumento (SensorML) o la determinación de un servicio web REST (WMS, WFS, WCS).

Paso 4: Generar XML

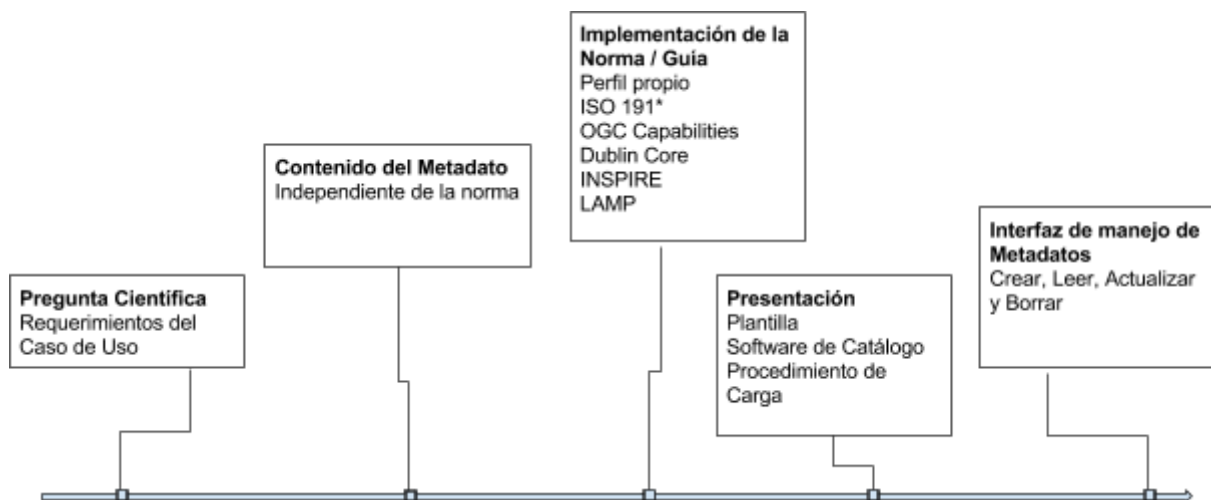
Muchas de estas preguntas, generadas en el punto 1, pueden ser consideradas como casos de uso de la documentación del metadato. Los casos de uso expresan las necesidades o requerimientos de los usuarios y describen cómo el sistema se dirige a esas necesidades. Documentan cómo se formulan las preguntas y, lo más importante, cómo el sistema presenta respuestas a estas preguntas. Una vez que se conocen las representaciones estándar, pasamos a convertir estos elementos en un producto concreto que nos permita de forma ordenada y dinámica implementar una manera única de representar el metadato de los productos que estamos analizando.

Cómo utilizamos como norma de marco la ISO 19115, la mejor implementación es la especificación técnica de la ISO 19139. La facilidad que nos permite es que ya está optimizada para poder utilizar el lenguaje XML y todas sus herramientas que permiten la interoperabilidad y la flexibilidad de estructura e implementaciones tecnológica.

	Tipo de documento: Documento teórico	Versión: 1.0
	Grupo de trabajo: Metadatos	Fecha : 15/11/2017
	Guía de Buenas Prácticas de Metadatos	Página 8 de 17

Paso 5: Gestionar la publicación


Llegando a la culminación de la publicación y gestión del metadato, nos encontramos con el desafío de que la herramienta planteada posea la capacidad de adaptación con respecto a la incorporación de estándares, la mejora continua del perfil utilizado para este tipo de dato y la propiedad de que la interfaz sea amigable y posea la metodología ABM.



Consistencia de geoservicios para evitar la duplicidad de información espacial y metadatos

En virtud de que, la cantidad de servidores de mapas tiene un crecimiento progresivo y se espera que siga esta tendencia, cada vez son más las organizaciones que publican su información geoespacial y la combinan con la de otras instituciones. Al consumir diversos servicios de mapas es posible que una misma capa sea proporcionada por más de un servicio. Por ejemplo, algunos organismos incorporan en sus servicios capas que provienen de otros servicios (publicación en cascada). Esto impide una clara y ágil integración de la información, es por ello que se propone estandarizar los servicios de mapas en cuanto a la organización y denominación de las capas ofrecidas en los servicios OGC de cada organismo.

Las recomendaciones para esta práctica se encuentran detalladas en el documento

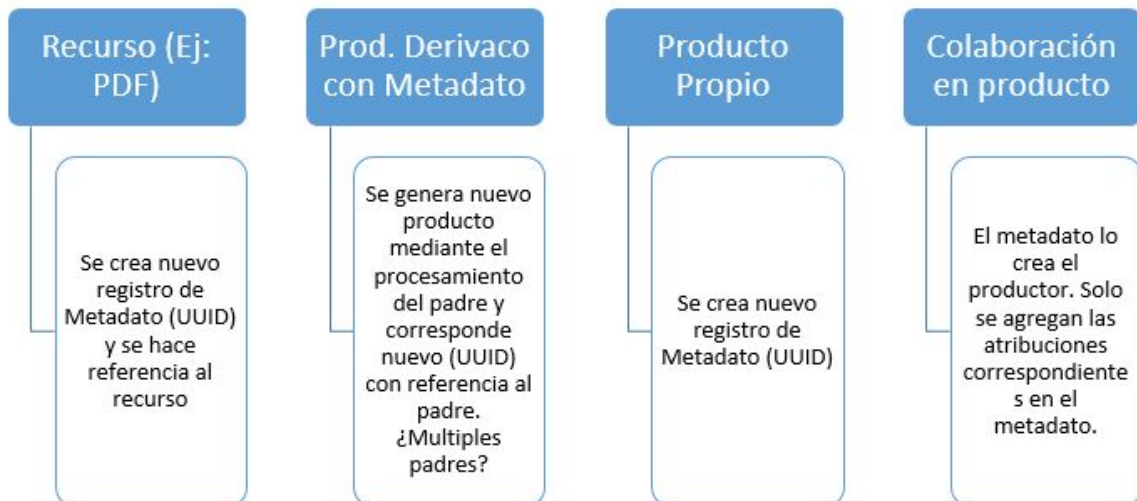
	Tipo de documento: Documento teórico	Versión: 1.0
	Grupo de trabajo: Metadatos	Fecha : 15/11/2017
	Guía de Buenas Prácticas de Metadatos	


[Normalización de Capas para Servicios OGC](#) producido por el grupo de trabajo Tecnología y Desarrollo de IDERA

Implementación de jerarquía y agrupamiento de fuentes de datos, únicas y múltiples

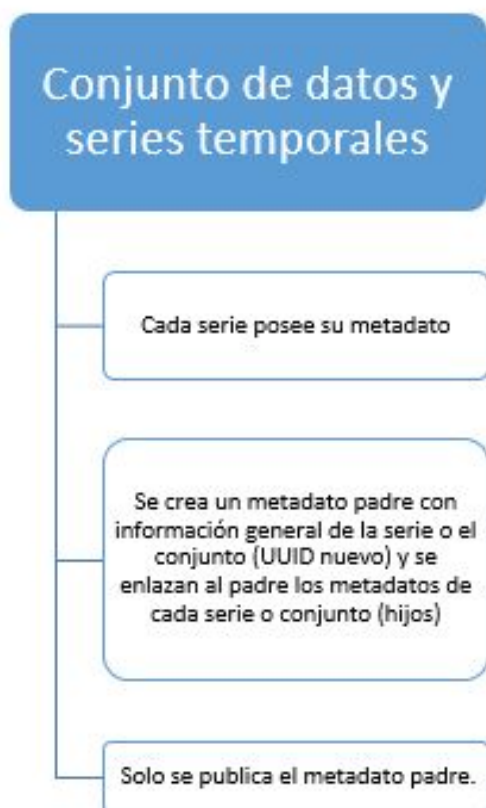
Al momento de comenzar con el diseño de los metadatos en una organización, tenemos que analizar la posibilidad de jerarquizarlos o agruparlos, lo que nos permite mejorar la búsqueda y visualización de los mismos. Los esquemas sugeridos por ISO 19115:2014 nos indican como proceder con las jerarquías y los productos a representar con metadatos.

Esquema 1: Jerarquía



	Tipo de documento: Documento teórico	Versión: 1.0
	Grupo de trabajo: Metadatos	Fecha : 15/11/2017
	Guía de Buenas Prácticas de Metadatos	Página 10 de 17

Esquema 2: Agrupamiento




Tratamiento de productos derivados

Muchas veces trabajamos con información de fuentes primarias la cual, a partir de análisis y procesamientos, derivará en nuevos productos.

Para registrar dichos procesos se deberán documentar de acuerdo a las siguientes recomendaciones:

- Si el metadato de una fuente primaria está en formato pdf o no existe el UUID, y a su vez se genera un derivado, es conveniente crear un metadato que transcriba lo contenido en el archivo pdf y a su vez adjuntarlo para poder incluirlo como fuente dentro del elemento linaje del metadato del producto derivado a través de un link o referencia de ubicación. Se recomienda citar como punto de contacto de la fuente al

	Tipo de documento: Documento teórico	Versión: 1.0
	Grupo de trabajo: Metadatos	Fecha : 15/11/2017
	Guía de Buenas Prácticas de Metadatos	Página 11 de 17

generador de la información primaria, y como punto de contacto del metadato al creador del producto derivado.

- En el caso de que no existan metadatos de la fuente primaria en ningún tipo de formato se recomienda cargar como metadato del producto derivado a la fuente original de información. Se sugiere citar como punto de contacto de la fuente al generador de la información primaria, y como punto de contacto del metadato al creador del producto derivado.

Formato de intercambio de metadatos


Se recomienda la utilización del formato XML (Extensible Markup Language) para compartir metadatos de una manera flexible. También es importante que, cuando se genera un producto para descargar, que el datos y el metadato se descargen de forma conjunta en un archivo comprimido.

Restricciones de uso y licencias

Respecto al acceso y uso de los datos, es importante que los mismos estén correctamente declarados en el metadato para evitar usos indebidos y problemas legales. El grupo de trabajo Institucional de IDERA dispone de un documento que amplía la información respecto a la temática, para la definición de estos elementos en el documento [Política de Datos y Glosario IDERA \(Lineamientos para el acceso, difusión, uso e interoperabilidad de la Información Geoespacial\)](#).

Recomendaciones Generales

- Proveer URLs directas a los recursos del metadato: permite la fácil y rápida ubicación de los mismos, así como también permite a los sistemas que cosechan los metadatos, acceder de forma programática a los recursos, como por ejemplo: un link a la descarga del archivo de dato y no a una pagina que a su vez posee otro link de descarga del dato;

	Tipo de documento: Documento teórico	Versión: 1.0
	Grupo de trabajo: Metadatos	Fecha : 15/11/2017
	Guía de Buenas Prácticas de Metadatos	Página 12 de 17


- Proveer la estructura correcta en el metadato para describir el acceso a la descarga de un recurso: en la sección distribución del metadato, se debe seleccionar el protocolo adecuado para el acceso a la descarga del recurso. En el caso en que un conjunto de datos posea un geo-servicio creado para solo a los fines de proveer acceso al mismo, se puede describir el servicio en la sección distribución del mismo metadato. De lo contrario el servicio debe poseer su propio metadato;
- Publicar en el catálogo únicamente metadatos propios, de datos propios o de productos derivados;
- Escribir resúmenes y títulos descriptivos: tanto el resumen como el título deberían responder a las preguntas: Que? Cuando? Donde?, en el caso que el metadatos sea para un producto publicado en un geoservicio, que el título sea el mismo que el descrito en el título del producto;
- Las colecciones o series de información homogéneas necesitan metadato a nivel de miembro y metadato a nivel de colección;
- Especificar fechas tanto del metadato como así también de los datos referidos;
- Publicar el metadato únicamente 1 vez y actualizarlo, no crear duplicados;
- Utilizar plantillas personalizadas, para las colecciones de datos homogéneas, para que los metadatos de estos sean igual de homogéneos (por ej. datos censales de la misma fuente, distintas variables).

Dato vs Metadato

Comúnmente cuando queremos dar una definición al concepto de metadato, decimos que son “*datos acerca de otros datos*”, pero **¿qué es el dato?**

El **dato** es *el registro que se obtiene como resultado de observar cierta característica de interés de un individuo (persona, animal, cosa o entidad de naturaleza abstracta) que constituye el objeto de estudio*. Por lo tanto, el dato supone:

- Una **unidad de análisis**: un “individuo” que ha sido observado/medido con cierta característica de interés.

	Tipo de documento: Documento teórico	Versión: 1.0
	Grupo de trabajo: Metadatos	Fecha : 15/11/2017
	Guía de Buenas Prácticas de Metadatos	

- Una **variable**: la observación/medición se realiza mediante criterios e instrumentos previamente determinados.
- Un **valor**: el dato se materializa en el registro de la medición realizada.

Estas 3 características mencionadas conforman la **estructura tripartita del dato**.

El resultado de la combinación de estas 3 partes será una matriz de datos, la cual conformará la **base de datos** que reunirá la información de un determinado universo según un tiempo y espacio. La misma se presenta en una estructura matricial de filas y columnas, donde por convención en las columnas se ingresan los atributos (**variables**) y en las filas los registros (**unidad de análisis**) con sus respuestas (**valores**).

Podemos citar algunos ejemplos de bases de datos para poder entender estos conceptos en relación a la construcción de objetos geográficos:


Modelo vectorial:

- OG: Departamento

variables unidad de analisis	FNA	GNA	NAM	CDE	SAG	SUP (km2)
Entidad 1	COMUNA 001	COMUNA	001	007	DGEyC - CABA	17,4
Entidad 2	COMUNA 002	COMUNA	002	014	DGEyC - CABA	6,1
Entidad 3	COMUNA 003	COMUNA	003	021	DGEyC - CABA	6,4

Valor

Modelo raster:

	Tipo de documento: Documento teórico	Versión: 1.0
	Grupo de trabajo: Metadatos	Fecha : 15/11/2017
	Guía de Buenas Prácticas de Metadatos	

- Matriz de datos resultado de una clasificación de usos de suelo para un recorte espacial.

Valor

Coordenadas - y	1	1	1	0	2
	1	0	0	2	2
	1	0	2	2	2
	Coordenadas - x				


Cuando decimos que la **unidad de análisis** puede ser observada o medida nos referimos a que podemos determinar las propiedades o atributos de ese “individuo” en relación a un criterio previamente definido en una categorización. Dado el conjunto de **unidades**, un **valor** es algo que puede predicarse de una unidad, y una **variable** es el conjunto de valores que forman esa clasificación. Dada esta relación, es importante reconocer las diferencias en la estructura interna de cada variable para poder medirlas y someterlas a distintas operaciones.

Las variables se clasifican según sus datos en:

- **Cuantitativas:** Denotan una cantidad del individuo observado y sus datos se expresan en número. Ej.: Cantidad de habitantes por radio censal
- **Cualitativa:** Denotan una cualidad del individuo y sus datos se expresan como una categoría predefinida del atributo observado. Ej.: Nombre de la localidad por radio censal.

Para los valores que toma cada variable existen 4 tipos de **niveles de medición**:

- **Nivel nominal:** Los datos se expresan con categorías que únicamente permiten clasificar a los individuos, sin establecer ningún tipo de orden o jerarquía. Al clasificar, se agrupan las unidades de análisis en clases mutuamente excluyentes en

	Tipo de documento: Documento teórico	Versión: 1.0
	Grupo de trabajo: Metadatos	Fecha : 15/11/2017
	Guía de Buenas Prácticas de Metadatos	Página 15 de 17

base a sus semejanzas y diferencias. Ej.: *Orientación educativa por establecimiento nivel secundario: Bachiller, Comercial, Técnico, Otros.*


- **Nivel ordinal:** Los datos se expresan con categorías que además de clasificar a los individuos, permiten establecer un orden entre ellos, definiendo la posición relativa de los individuos con respecto a esta característica. Ej.: *Nivel de enseñanza por establecimiento educativo.*
- **Nivel intervalar:** Los datos se expresan solamente mediante números enteros. Ej.: *Cantidad de alumnos matriculados por establecimiento educativo.*
- **Nivel racional:** Los datos se expresan en números no enteros. Ej.: *Superficie en m² del establecimiento educativo.*

Los datos pueden provenir de diferentes fuentes u orígenes:

- **Datos primarios:** Son aquellos conocidos como “datos brutos”. Se encuentran en su forma original en que fueron registrados sin haber sido sujeto a algún tratamiento posterior. Pueden ser **propios**, aquellos que son recolectados por quien los produce, o **registros de terceros**, producidos por otros pero se ajustan a las necesidades de quien lo utiliza.
- **Datos secundarios:** Son aquellos que fueron producidos por terceros y que fueron sometidos a algún tratamiento posterior para su uso.

Según su etimología, la palabra “**dato**” proviene del latín “*datum*”: *lo que se da*. Sin embargo, esta definición es engañosa ya que el dato, no es nada dado sino que hay que producirlo y presupone toda una serie de operaciones para su construcción, implicando elecciones teóricas-metodológicas. El dato es producto de un acto clasificatorio y solo tiene sentido de información en el contexto en el cual fue enunciado.

Todo dato espacial, ya sea en formato vectorial o raster, es una representación de un modelo, una abstracción, una de las tantas “vistas” o “recortes” posibles de un universo. Si analizamos los ejemplos de base de datos citados anteriormente, en especial el del modelo ráster, con solo mirarlo nos daríamos cuenta que necesitamos la información de cómo se construyó el dato y que significan cada uno de los valores que adopta, o el sistema de referencia espacial ya que por sí mismo no nos dice nada.

	Tipo de documento: Documento teórico	Versión: 1.0
	Grupo de trabajo: Metadatos	Fecha : 15/11/2017
	Guía de Buenas Prácticas de Metadatos	Página 16 de 17

Por lo tanto, es fundamental registrar el proceso de construcción del dato para poder entender los mismos a la hora de hacer uso de éstos. Este registro es lo que llamamos **“metadato”**, *un conjunto de atributos o elementos necesarios para describir y documentar un recurso en particular.*


El metadato asegura que el uso del dato no sea equívoco, permitiendo al productor del mismo la descripción del recurso para que otros usuarios puedan entender los supuestos y limitaciones y así evaluar las posibles aplicaciones del mismo para sus intenciones de uso.

Según la norma ISO 19115:2014, el alcance del metadato es proveer información de los datos y servicios geográficos producidos sobre:

- Identificación
- Extensión
- Calidad
- Aspectos temporales y espaciales
- Contenido
- Referencias espaciales
- Representación
- Distribución

Referencias bibliográficas

- Aguirre Cesar, Niño M. Fernanda, Simonetti Eduardo (2005) Manual de Estadística Aplicada en las Ciencias Sociales y Humanas. Editorial Universitaria. Universidad Nacional de Misiones
- Denis Baranger (2009) “Construcción y análisis de datos. Introducción al uso de técnicas cuantitativas en la investigación social”. Editorial Posadas.
- Documento de Esquema de Metadatos de IDERA. Versión: 1.0. Documento técnico elaborado por el grupo de trabajo Metadatos de IDERA
- IDERA (2016) “Normalización de capas para servicios OGC”

	Tipo de documento: Documento teórico	Versión: 1.0
	Grupo de trabajo: Metadatos	Fecha : 15/11/2017
	Guía de Buenas Prácticas de Metadatos	Página 17 de 17

- ISO 19115 Geographic Information - Metadata. 2014
- https://geo-ide.noaa.gov/wiki/index.php?title=Use_Cases_to_CRUD