



# SEGMENTOS DE INFORMACIÓN

Catálogo de Metadatos.

Cuarto Plan de Acción Nacional de Gobierno Abierto

Fortalecimiento de las unidades de informática y tecnología para desarrollar y optimizar las plataformas tecnológicas institucionales interoperables del sector seguridad





## COMPROMISO 6, HITO 10

Dirección de Monitoreo y Comunicación

# CATÁLOGO DE METADATOS

## PLATAFORMA TECNOLÓGICA INTEROPERABLE DEL SISTEMA NACIONAL DE SEGURIDAD

Al publicar conjuntos de datos, los cargadores o mantenedores generalmente proporcionan metadatos complementarios. Los metadatos describen información importante sobre el conjunto de datos, como título, licencia, cuerpo de publicación, frecuencia de actualización, etc. En ese sentido la Secretaría Técnica del Consejo Nacional de Seguridad toma como una necesidad para el proceso de interoperabilidad entre sistemas de información, diseñar un catálogo de metadatos en materia de Seguridad de la Nación con esto se busca basarse en estándares internacionales vigentes.

### Elementos conceptuales

- a. Los **metadatos** son datos que describen otros datos. En general, un grupo de metadatos se refiere a un grupo de datos que describen el contenido informativo de un objeto al que se le denomina recurso. El concepto de metadatos es análogo al uso de índices para localizar objetos.
- b. La **interoperabilidad** es una característica de los sistemas, que refiere a su capacidad para interactuar y funcionar con otros productos o sistemas, de manera compatible. Implica que las instituciones involucradas compartan información y conocimiento, mediante el intercambio de datos entre sus respectivos sistemas de tecnología de la información y comunicación.
- c. La **interoperabilidad semántica**<sup>1</sup> garantiza que el formato y el significado exacto de la información intercambiada se comprendan y conserven en todos los intercambios entre las partes, es decir, "que lo que se transmite se entiende". En el contexto del Fondo Europeo de Inversiones, la interoperabilidad semántica engloba los siguientes aspectos semánticos y sintácticos
  1. El aspecto semántico se refiere al significado de los elementos de los datos y la relación entre ellos, incluye la creación de vocabularios y esquemas para describir los intercambios de los datos y garantiza que todas las partes que se comunican entienden de la misma manera los elementos de datos;
  2. El aspecto sintáctico se refiere a la descripción del formato exacto de la información que se va a intercambiar en términos de gramática y formato.

### Estándares de metadatos<sup>2</sup>

- a. **Data Catalog Vocabulary (DCAT)**, una recomendación del W3C establecida el 16 de enero de 2014, está diseñada para "facilitar la interoperabilidad entre los catálogos de datos publicados en la Web" (W3C). El objetivo principal de DCAT es mejorar la interoperabilidad de los catálogos de datos y hacer que las aplicaciones consuman fácilmente metadatos de múltiples catálogos.
- b. **Esquema de metadatos de descripción de activos (ADMS)**, es un esquema de metadatos creado por el Programa de Soluciones de Interoperabilidad de la UE para las Administraciones Públicas Europeas

<sup>1</sup> [https://eur-lex.europa.eu/resource.html?uri=cellar:2c2f2554-0faf-11e7-8a35-01aa75ed71a1.0010.02/DOC\\_3&format=PDF](https://eur-lex.europa.eu/resource.html?uri=cellar:2c2f2554-0faf-11e7-8a35-01aa75ed71a1.0010.02/DOC_3&format=PDF)

<sup>2</sup> <http://knowhow.opendatamonitor.eu/odresearch/metadata-standards/>



## COMPROMISO 6, HITO 10

### Dirección de Monitoreo y Comunicación

- (ISA). El objetivo de ADMS es ayudar a los editores de estándares a documentar los metadatos de los estándares, como nombre, estado, tema, versión, etc.
- c. **El Perfil de aplicación DCAT (DCAT-AP)** para portales de datos en Europa es una especificación que reutiliza los términos de DCAT, ADMS, etc., y agrega más especificidad al identificar elementos obligatorios, recomendados y opcionales que se utilizarán para un determinado dato abierto. catalogar. Estudios realizados por la comisión de la UE (Vickery, 2011) han demostrado que las empresas y los ciudadanos enfrentan dificultades para buscar y reutilizar conjuntos de datos del sector público. Por lo tanto, la disponibilidad de un método unificado para describir conjuntos de datos en un formato legible por máquina con un pequeño número de metadatos comúnmente acordados podría mejorar en gran medida la correferencia e interoperabilidad entre los diferentes catálogos de datos. DCAT-AP se desarrolla en este contexto y se espera que se aplique en los portales de datos abiertos en los países de la UE.
  - d. **CKAN** es el software de portal de datos abierto más utilizado hasta la fecha y, como tal, su esquema de metadatos respectivo es muy relevante para el proyecto ODM. A diferencia de otros estándares del W3C mencionados anteriormente, los metadatos de CKAN se exponen a través de la API RESTful y los cargadores de datos deberán completar los metadatos con la solicitud de la API.
  - e. **INSPIRE** es una Directiva del Parlamento Europeo y del Consejo cuyo objetivo es establecer una "infraestructura de datos espaciales en toda la UE para dar acceso a la información que puede utilizarse para apoyar las políticas ambientales de la UE en diferentes países y sectores públicos" (INSPIRE). El alcance real de esta información corresponde a 34 temas ambientales, cubriendo áreas que tienen relevancia intersectorial, por ejemplo, direcciones, edificios, distribución de la población y demografía.
  - f. **El Common Core Metadata Schema (CCMS)**, se basa en DCAT y proporciona vocabulario mutuo al que pueden mapearse diferentes esquemas de metadatos de datos abiertos. El estándar consta de una serie de esquemas (términos de vocabulario jerárquico) que representan las cosas que con mayor frecuencia se buscan en la web. CCMS también proporciona las asignaciones a sus equivalentes en otros estándares.
  - g. **VOID** es un "vocabulario de esquema RDF para describir metadatos sobre conjuntos de datos RDF" (VOID). Su propósito principal es cerrar la brecha entre los editores de datos y los consumidores de datos utilizando un vocabulario exclusivo para describir diferentes atributos del conjunto de datos. Los conceptos centrales relacionados con los conjuntos de datos abiertos son: vacío: conjunto de datos, vacío: conjunto de enlaces, vacío: subconjunto.
  - h. **Schema.org** es una colección de esquemas (en formato RDF / Microdata) que los webmasters pueden usar para marcar páginas HTML de formas reconocidas por los principales motores de búsqueda. Schema.org cubre muchos dominios y hay clases y propiedades definidas como DataCatalog y Dataset. El recolector de metadatos dentro del proyecto ODM puede utilizar el vocabulario de schema.org para descubrir los conjuntos de datos y los catálogos de datos alojados en un determinado sitio web.
  - i. **Google Dataset Publishing Language** es un "lenguaje de representación para los datos y metadatos de conjuntos de datos". Los conjuntos de datos descritos con este formato se pueden visualizar directamente desde Google Public Data Explorer.

## Propuesta

De acuerdo a lo anterior se ha seleccionado el estándar del Vocabulario de Catálogo de Datos (DCAT) para generar una propuesta inicial que rija un catálogo de metadatos ya que una de las plataformas más amplias por sus extensiones y con mayor comunidad trabajando en su estandarización e implementación bajo



## COMPROMISO 6, HITO 10

*Dirección de Monitoreo y Comunicación*

iniciativas europeas como red.es o bien iniciativa “aporta” de España. Se contempla DCAT como un referente que puede coadyuvar a la definición del Catálogo de Metadatos en materia de Seguridad de la Nación.

Al usar DCAT para describir los conjuntos de datos (datasets) que se almacenan, los editores facilitan su descubrimiento y permiten que otras aplicaciones puedan consumir y manipular más fácilmente los metadatos albergados en diversos catálogos. Además, facilita la búsqueda federada de conjuntos de datos a través de una plataforma integrada. Estos metadatos descritos mediante DCAT pueden servir como archivo para facilitar la preservación digital.

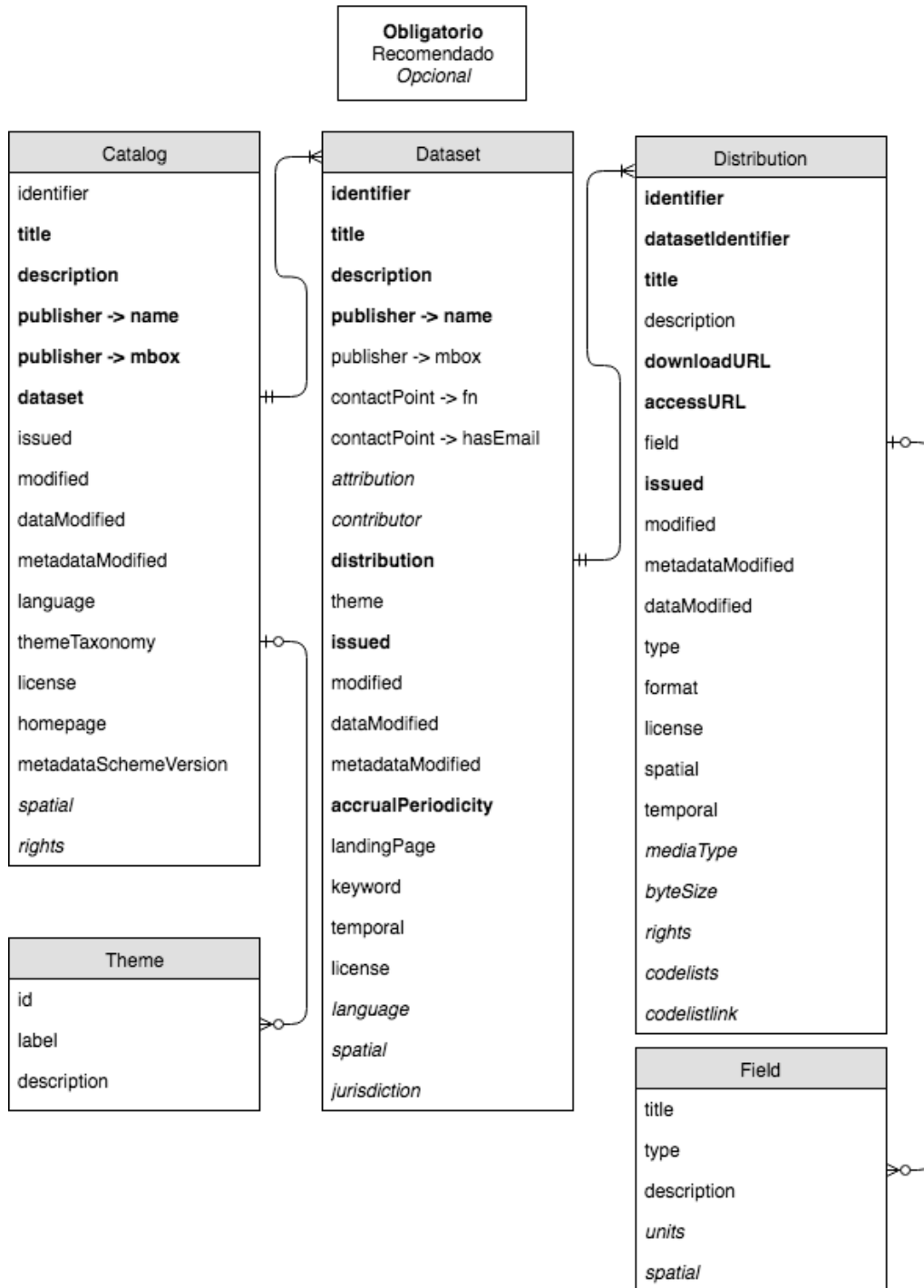
Al construir un catálogo de metadatos se busca una base común entre las instituciones del Sistema Nacional de Guatemala que pueda ser una referencia y facilite las iniciativas de interoperabilidad entre conjuntos de datos que se generan en temas de seguridad, difundiendo soluciones comunes entre los mismos. En ese sentido es importante recalcar los principios que servirán de guía para la implementación del catálogo:

- a. **Se basa en la aplicación europea de DCAT (DCAT-AP).** Toma como base el perfil de metadatos de catálogos de datos abiertos de más amplia difusión, y lo extiende.
- b. **Sigue la misma filosofía de extensibilidad que DCAT**
  1. Preserva el uso de todos los campos obligatorios
  2. Sin un campo no obligatorio está definido en el perfil base, es preferible utilizar ese antes de crear uno nuevo.
  3. Crea nuevos campos de metadatos, que resuelven casos de uso muy específicos del Sistema Nacional de Seguridad, no previstos por DCAT-AP.
- c. **Reutiliza soluciones de uso por países que han implementado estos estándares.** Agrega campos de metadatos adicionales en base a la experiencia de su uso en otros países.
- d. **Define nombres de variables en inglés y en *lowerCamelCase*,** en sintonía con los nombres del perfil DCAT.



**COMPROMISO 6, HITO 10**

Dirección de Monitoreo y Comunicación





## COMPROMISO 6, HITO 10

Dirección de Monitoreo y Comunicación

Nombre	Descripción	Ejemplo	Variable	Tipo
Identificador	Identificador único del dataset, este identificador debe ser único para todo el catálogo.	"dataset-ejemplo-35782"	identifier	String
Título	Nombre asignado al dataset tal como será publicado. Debe ser claro y lo suficientemente abstracto como para abarcar la multiplicidad de distribuciones que contiene. Se recomienda no exceder los 100 caracteres en la mayoría de los casos.	NNA migrantes no acompañados	title	String
Descripción	Descripción del contenido del dataset de un modo claro y conciso. Se recomienda no exceder los 500 caracteres en la mayoría de los casos.	Dat	description	String
Autor	Autoridad responsable de la publicación del dataset. Se recomienda adoptar un vocabulario controlado con las dependencias del país de que se trate y definir una forma de expresar la pertenencia jerárquica de la organización (Ej.: utilizar puntos o guiones medios para separar niveles jerárquicos, expresarlos en forma descendente o ascendente, etc.)	Procuraduría General de la Nación	publisher -> name	String
Correo electrónico del autor	Correo electrónico de contacto de la autoridad responsable de la publicación del dataset.	<a href="mailto:uip@pgn.gob.gt">uip@pgn.gob.gt</a>	publisher -> mbox	String
Área/Persona de contacto	Área/persona de contacto que puede brindar información relevante sobre el dataset. Este es un punto de contacto más próximo al mantenimiento u operación de este dataset.	Procuraduría General de la Nación Unidad de Información Pública	contactPoint -> fn	String
Correo electrónico del	Correo electrónico del área/persona de contacto que puede brindar información relevante sobre el dataset.	<a href="mailto:uip@pgn.gob.gt">uip@pgn.gob.gt</a>	contactPoint -> hasEmail	String



## COMPROMISO 6, HITO 10

Dirección de Monitoreo y Comunicación

área/persona de contacto				
<b>Fuente primaria</b>	Fuente original o primaria de los datos publicados en el dataset. Se utiliza cuando la entidad responsable de la publicación del dataset, no es la entidad que produce los datos.	Instituto Guatemalteco de Migración	attribution	String
<b>Colaboradores</b>	Entidades, organizaciones o personas adicionales que son o fueron responsables de recolectar, crear o contribuir en forma alguna al desarrollo del activo de datos. Los valores deben separarse por comas.	Secretaría de Bienestar Social	contributor	String
<b>Distribuciones</b>	Lista de distribuciones que pertenecen al dataset y sus metadatos. Cada distribución se representa con un objeto ("{}") donde se describen los metadatos especificados para la clase "distribution" de este perfil de metadatos.	[{...}, {...}]	distribution	Array
<b>Temática(s)</b>	Temática/s o categoría/s a la/s que se refiere el dataset al ser publicado. Un dataset puede pertenecer a más de una categoría temática, de manera que el tipo de valor de este campo es una lista de categorías. La taxonomía a utilizar debe ser creada por la autoridad responsable del catálogo.	["seguridad", "exterior"]	theme	Array
<b>Fecha de publicación</b>	Fecha de publicación del dataset. Según el formato ISO-8601, tipeado como fecha simple o fecha con hora, con el mayor detalle posible que sea relevante para el dataset.	"2016-04-14T19:48:05.433640" para especificar fecha y hora "2016-04-14" para especificar fecha únicamente	issued	String
<b>Fecha de última actualización/modificación</b>	Fecha de última actualización/modificación de dataset (ya sea de sus datos o de sus metadatos). Según el formato ISO-8601, tipeado como fecha simple o fecha con hora, con el mayor detalle posible que sea relevante para el dataset.	"2016-04-19T19:48:05.433640" para especificar fecha y hora "2016-04-19" para especificar fecha únicamente	modified	String





## COMPROMISO 6, HITO 10

Dirección de Monitoreo y Comunicación

<b>Fecha de última actualización/modificación</b>	Fecha de última actualización/modificación de los datos del dataset. Según el formato ISO-8601, tipeado como fecha simple o fecha con hora, con el mayor detalle posible que sea relevante para el dataset.	"2019-04-19T19:48:05.433640" para especificar fecha y hora "2019-04-19" para especificar fecha únicamente	dataModified	String
<b>Fecha de última actualización/modificación</b>	Fecha de última actualización/modificación de los metadatos del dataset. Según el formato ISO-8601, tipeado como fecha simple o fecha con hora, con el mayor detalle posible que sea relevante para el dataset.	"2019-04-19T19:48:05.433640" para especificar fecha y hora "2019-04-19" para especificar fecha únicamente	metadataModified	String
<b>Frecuencia de actualización</b>	Frecuencia con la que se actualiza el dataset. Se recomienda especificar períodos normalizados con formato ISO-8601, agregando el campo "eventual" para datasets que se publican con una frecuencia eventual o no especificada.	"R/P1Y" para datasets que se actualizan anualmente	accrualPeriodicity	String
<b>Etiqueta(s)</b>	Palabras que describen el título o el contenido del recurso. Es necesario que las etiquetas se encuentren correctamente escritas, en plural y respetando la existencia de tags anteriores. Etiquetas que colaboran en la búsqueda de los usuarios. Cuanto más amplia y uniforme sea la lista de tags mayor será su eficiencia.	["NNA", "migrantes", "deportados"]	keyword	Array
<b>Cobertura temporal</b>	Período de tiempo cubierto por el dataset. El intervalo de tiempo está formado por una fecha de inicio y una de fin separadas por "/", en formato ISO 8601, con el nivel de especificidad requerido por el dataset. Opcionalmente, se puede especificar una sola fecha (en lugar de dos) y esta se interpretará como la fecha de inicio, mientras que el final permanece no especificado.	2015-01-01/2015-12-31 2015-01-01T00:45:00Z/2016-01-15T00:06:00Z 2015-01-01 2015-01-01T00:45:00Z	temporal	String
<b>Licencia</b>	Indica la licencia bajo la cual el dataset y todas sus distribuciones están disponibles mediante un enlace a la licencia o documento de la licencia seleccionada, o mediante el título textual de la licencia tal	"http://opendatacommons.org/licenses/dbcl/1-0/" si se utiliza un enlace "Open Database License (ODbL) v1.0" si se consigna el nombre de la licencia a utilizar	license	String





**COMPROMISO 6, HITO 10**

*Dirección de Monitoreo y Comunicación*

como aparece en la lista de <http://opendefinition.org/licenses/> . Se recomienda utilizar la licencia "Open Database License (ODbL) v1.0". Un dataset hereda por default la licencia general del catálogo salvo que se especifique una licencia diferente en este campo. Las distribuciones del dataset heredan esta licencia salvo que especifiquen una diferente.

<b>Idioma(s)</b>	<p>Lenguaje para la descripción de los metadatos del dataset. Hay 2 estándares ISO que pueden ser utilizados para este campo: (a) ISO 639-1 (2 letras) (b) ISO 639-2/T (3 letras) es el más recomendado. Puede definirse 1 o más lenguajes en una lista. Los lenguajes especificados para un dataset, sobreescriben a los del catálogo. Si este campo está vacío el dataset hereda los lenguajes del catálogo. (Link a los estándares ISO: <a href="https://www.loc.gov/standards/iso639-2/php/code_list.php">https://www.loc.gov/standards/iso639-2/php/code_list.php</a>)</p>	<p>["es"] para un lenguaje ISO 639-1 ["spa", "eng"] para dos lenguajes ISO 639-2</p>	language	Array
<b>Cobertura geográfica</b>	<p>Una región o lugar determinado al que el dataset haga referencia. Esta se puede expresar mediante diferentes estándares. Se recomienda el desarrollo de una guía de unidades territoriales y político-administrativas específica para el país que administra el catálogo, donde se especifiquen los códigos oficiales que deben usarse para señalar cobertura geográfica. Una región o un lugar determinado. Puede tomar valores: a) de países, provincias, departamentos, condados, municipios o localidades, según las recomendaciones de la guía de unidades territoriales o político-administrativas oficial del país que administra el catálogo. b) un área de coordenadas representada por latitud/longitud en el orden: mínima longitud,</p>	<p>"GUA" es el código para la República de Guatemala. "0719" es el código asignado por INE al municipio de Santiago Atitlán [-58.111111, -35.111111, -57.111111, -33.111111] es un bounding box [-58.111111, -35.111111] es un punto geográfico "http://sws.geonames.org/6255146"</p>	spatial	Array or String





## COMPROMISO 6, HITO 10

*Dirección de Monitoreo y Comunicación*

minima latitud, maxima longitud, maxima latitud. c) un punto geográfico representado por latitud/longitud.

**Jerarquía,  
orden o nivel  
jurisdiccional**

Clasificación del dataset por su jerarquía, orden o nivel jurisdiccional. Indica el tipo y alcance de la jurisdicción legal que está detrás de la publicación del dataset. El valor puede ser "national" o "territorial" (cubriendo el caso nacional por un lado y todas las jurisdicciones sub-nacionales por otro), o pertenecer a un vocabulario controlado alternativo definido por el país o la organización responsable de la política de apertura (Ej.: "federal", "provincial" o "municipal").

"national" "territorial"





SECRETARÍA TÉCNICA DEL  
CONSEJO NACIONAL DE SEGURIDAD  
REPÚBLICA DE GUATEMALA