

ESTÁNDAR DE DATOS PARA CONJUNTOS DE DATOS DE PROYECTOS DE INVESTIGACIÓN

FECHA DE ACTUALIZACIÓN: JUNIO 2019

Descripción del estándar

La adopción de estándares para la publicación de conjuntos de datos permite la interoperabilidad de los datos por sistemas informáticos, además de facilitar el uso de los datos para nuevos análisis y generar nuevo conocimiento.

Para la definición de los términos adecuados para describir los conjuntos de datos de proyectos de investigación, se revisaron los esquemas de metadatos utilizados por dos de las principales plataformas en actual uso para la publicación de este tipo de recursos: la plataforma CKAN utiliza el esquema Data Catalog Vocabulary (DCAT, <https://www.w3.org/TR/vocab-dcat/>) y la plataforma de datos de investigación INVENIO basa su diseño de metadatos en DataCite (<https://datacite.org/index.html>).

También se revisaron los vocabularios controlados desarrollados por Schema.org, (<https://schema.org/>) que permiten la mejor recuperación de los conjuntos de datos y la descripción de sus contenidos.

Considerando estas fuentes y las necesidades propias de la Plataforma de Datos Abiertos de Investigación, se diseñó el siguiente esquema de metadatos.

Diccionario

Termino	Definición
id	Número consecutivo e irrepitable asignado por la plataforma de Datos Abiertos de Investigación.
guid	Identificador construido por el código de la institución, el código de la comunidad y el identificador numérico en la base de datos.
identifier	Identificador interno de la plataforma.



ESTÁNDAR DE DATOS PARA CONJUNTOS DE DATOS DE PROYECTOS DE INVESTIGACIÓN

FECHA DE ACTUALIZACIÓN: JUNIO 2019

Termino	Definición
alternate_identifiers	Lista de identificadores del recurso en otras plataformas, debe indicarse la cadena del indicador y la plataforma en la que el recurso se identifica.
communities	Nombre de la comunidad a la que pertenece el recurso, nombre proviene de las comunidades registradas en la Plataforma de Datos Abiertos de Investigación.
title	Nombre o título del recurso, deber ser un nombre que describa de forma precisa y consisa el tipo de recurso, la temática y la procedencia.
version	Número que identifica la versión del recuso. Se sugiere utilizar cadenas numéricas y seguir algún sistema de versionado https://semver.org
resource_type	Descripción del recurso: base de datos, código, texto, etc.
database	Características técnicas de la base de datos: formato de la base de datos, lenguaje, número de columnas, número de filas, estándar de datos utilizado, última fecha de actualización de los registros.
creators	Nombre del principal investigador involucrado en la producción del recurso.
contributors	Nombre de las personas o instituciones responsables de la compilación, manejo, distribución o desarrollo del recurso.
date	Lista de las fechas y el tipo de fecha relacionadas con el recurso. Ejemplos: creación: 2019-05-29, embargo: 2019-05-29, actualización: 2019-05-29.



ESTÁNDAR DE DATOS PARA CONJUNTOS DE DATOS DE PROYECTOS DE INVESTIGACIÓN

FECHA DE ACTUALIZACIÓN: JUNIO 2019

Termino	Definición
embargo_date	Fecha después de la cual el recurso puede ser público.
files	Descripción técnica de los archivos: nombre del archivo, tipo MIME, tamaño.
subjects	Temas, palabras claves, clasificación o frases que describen al recurso.
description	Información adicional que no se encuentra en ninguna de las categorías y que describen el contenido y la parte técnica de la base de datos.
related_identifiers	Lista de identificadores relacionados con el recurso.
keywords	Palabras claves normalizadas.
language	El idioma principal del contenido del recurso.
license	Licencia de datos del recurso.
references	Lista de referencias utilizadas por el recurso.
access_condition	Condiciones de acceso al recurso.
access_right	Tipo de acceso al recurso: acceso abierto, recurso con fecha de embargo, con acceso restringido, acceso cerrado.
notes	Notas técnicas para el procesamiento del recurso antes de su publicación.



ESTÁNDAR DE DATOS PARA CONJUNTOS DE DATOS DE PROYECTOS DE INVESTIGACIÓN

FECHA DE ACTUALIZACIÓN: JUNIO 2019

Termino	Definición
modified	Fecha de la última modificación del registro en la base de datos de origen.
_creation	Fecha de creación del registro en la Plataforma de la DGRU.
_creation_user	Nombre de usuario que creo el registro en la plataforma, en caso de no contar con el dato se utilizará el código del dataset_id.
_last_modified	Fecha de última modificación del registro dentro de la Plataforma.
_last_modified_user	Nombre de usuario que realizó la última modificación del registro dentro de la Plataforma de la DGRU.
responsable_quality_control	Nombre de usuario responsable del proceso de Control de Calidad Estructural en la DGRU.