



Datasets: plantilla normalizada para la descripción de registros en Digital.CSIC

Oficina Técnica de Digital.CSIC

01/12/2017

La descripción recomendada para datos de investigación debe ser en inglés si es posible e incluye los siguientes aspectos:

- Contexto, descripción del proyecto y propósito de la investigación, metodología utilizada
- Naturaleza de los datos, historia de los datos, contenido y estructura, terminología, software, fecha de creación y fechas de modificación, versiones, responsables y participantes
- Formatos de ficheros, estructura y nomenclatura de los ficheros
- Aspectos legales, políticas de acceso y seguridad

La descripción de los datos debe ser lo suficientemente completa como para poder responder las siguientes preguntas, si no, su utilidad para otros investigadores será relativa:

- *¿Quiénes han producido los datos?*
- *¿Es el título del dataset suficientemente específico?*
- *¿Por qué han sido creados los datos?*
- *¿Qué limitaciones tienen los datos (por ejemplo, datos confidenciales que han sido eliminados)?*
- *¿Cómo deben interpretarse los datos?*
- *¿Hay lagunas en los datos o dan una visión completa del tema estudiado?*
- *¿Qué procesos han generado los datos?*
- *¿Qué miden los datos en las columnas de los ficheros?*
- *¿Qué software es necesario para poder leer los datos?*
- *¿Cómo deben citarse los datos?*
- *¿Pueden reutilizarse los datos? ¿Qué licencia de uso tienen asignada?*
- *¿Existen más versiones de los datos? ¿Dónde?*
- *¿Se han definido los términos técnicos y acrónimos a los que hacen referencia los datos?*
- *¿Se han cualificado los parámetros geográficos y cronológicos de los datos?*
- *¿Las palabras clave son suficientemente específicas a los datos? ¿Se basan en algún tesoro?*
- *¿Cómo se llama el proyecto de investigación en que se encuadran los datos?*
- *¿Quién ha financiado la producción y la gestión de los datos?*

Además, la política de datos de Digital.CSIC recomienda a sus autores la generación de un fichero txt *read.me* con más información (ver más detalles en la última página de la plantilla).

Se recomienda ver [Buenas prácticas y política de datos de investigación de Digital.CSIC](#) antes de describir y depositar un conjunto de datos en el repositorio.

DESCRIPTOR	METADATO DUBLIN CORE	CUALIFICADOR	CARÁCTER
AUTOR	dc.contributor	Author	Obligatorio
<p>Nombres de los autores de los datos</p> <p>Deben citarse en campos independientes todos los autores en el orden en que aparecen. En la generación y gestión de datos científicos pueden estar involucrados distintos tipos de autores y todos deben ser reconocidos en el registro, por ejemplo: creador de los datos, analista de los datos, gestor de los datos, gestor del proyecto de datos, supervisor de los datos etcétera.</p> <p>Se recomienda hacer una búsqueda previa para comprobar si un autor ya ha sido introducido en el índice de autores para citarlo de la misma manera y evitar duplicidades https://digital.csic.es/browse?type=author</p>			



TÍTULOS

TÍTULO	dc.title		Obligatorio
<p>Nombre del conjunto de datos</p> <p>Es muy importante asignar un título unívoco al dataset que se está describiendo y evitar, por ejemplo, asignarle el mismo título que el del artículo científico con el que se asocia, o con el proyecto al que pertenece.</p> <p>Se escribe en minúsculas, aunque los sustantivos pueden aparecer con la inicial en mayúsculas si figura así en el original.</p> <p>Evitar el punto y final.</p> <p>Si se trata de una de las versiones disponibles del conjunto de datos, se debe indicar entre paréntesis (por ejemplo, v.2.2).</p>			

OTROS TÍTULOS	dc.title	Alternative	Opcional
<p>En caso de que haya variantes de título (también a nivel de idioma).</p>			

FECHAS

DESCRIPTOR	METADATO DUBLIN CORE	CUALIFICADOR	CARÁCTER
FECHA DE PUBLICACIÓN	dc.date	issued	Obligatorio
<p>Fecha de publicación o de distribución pública</p> <p>Como mínimo escribiremos el año.</p> <p>Si no ha sido publicado podemos poner la fecha del depósito.</p> <p>Si terminamos de subir el registro y hubiera un error en la fecha habrá que corregirlo en modo edición y en este caso el formato sería aaaa-mm-dd</p>			

FECHA DE INICIO	dc.date	created	Recomendado
<p>Fecha de creación del conjunto de datos</p> <p>Formato: aaaa-mm-dd</p> <p>Este metadato es muy recomendable ya que en general existe un lapso de tiempo entre la creación de los datos y su publicación en el repositorio. Esta fecha sirve, además, para determinar a partir de cuándo el dataset está protegido por copyright.</p> <p>Disponible solo en el modo de edición.</p>			



FECHA DE EMBARGO	dc.embargo	terms	Opcional
<p>Fecha de embargo hasta hacer público el dataset</p> <p>Formato: aaaa-mm-dd</p> <p>Hay que adjuntar el fichero válido.</p> <p>Para embargar un trabajo consulte la siguiente guía: Guía para embargar registros en Digital.CSIC</p> <p>Una vez que el dataset está embargado en el repositorio, y durante todo el periodo de carencia, se activa la funcionalidad de “<i>Petición de copia privada</i>” por la que, por ejemplo, pueden solicitar una copia del mismo los revisores del artículo asociado a los datos.</p>			

FECHA DE ACTUALIZACIÓN	dc.date	updated	Opcional
<p>Fecha en que se ha actualizado el registro</p> <p>Formato: aaaa-mm-dd</p> <p>Disponible solo en el modo de edición.</p>			

MARCO TEMPORAL Y CRONOLÓGICO

DESCRIPTOR	METADATO DUBLIN CORE	CUALIFICADOR	CARÁCTER
UBICACIÓN GEOGRÁFICA	dc.coverage	spatial	Opcional
<p>Ubicación física, cobertura geográfica</p> <p>Hace referencia al lugar de recogida de datos o al lugar al que hacen referencia los datos.</p> <p>Es muy recomendable, ya que es una información contextual significativa para muchos datasets y la base para realizar análisis y visualizaciones de geolocalización.</p> <p>Se recomienda consultar las entradas normalizadas de vocabularios controlados como Getty Thesaurus of Geographic Names o el más genérico Geonames y repetir el metadato con la URL de la base de datos consultada.</p> <p>Recomendado usar las entradas en inglés.</p> <p>Metadato repetible tantas veces cuantos lugares haya que referenciar.</p> <p>Es también recomendable indicar la longitud y la latitud en su forma decimal.</p> <p>Por ejemplo: dc.coverage.spatial: name=Chile; lat=-30.0000; long=-71.0000 dc.coverage.spatial: http://www.geonames.org/3895114/chile.html</p>			



COBERTURA TEMPORAL	dc.coverage	temporal	Opcional
<p>Cobertura temporal</p> <p>Hace referencia al periodo que cubre el dataset o el periodo de recogida de datos.</p> <p>Es un metadato muy recomendable por la información contextual significativa que aporta.</p> <p>Incluir siempre las fechas con el formato aaaa-mm-dd, precedido por “start” o “end” según corresponda.</p> <p>Ejemplo:</p> <p>start=1981-01-31; end=2001-01-01</p>			

CITACIÓN

REFERENCIA/CITACIÓN	dc.identifier	citation	Obligatorio
<p>Referencia bibliográfica del conjunto de datos</p> <p>Aplicamos la recomendación de FORCE11:</p> <p>Authors; Year; Dataset title; Data repository or archive; Version (if any); Persistent identifier (e.g. DOI)</p> <p>Ejemplo:</p> <p>Tur, Cristina ; Traveset, Anna ; Eguíluz, Víctor M.; 2017; "Plant survival and keystone pollinator species in stochastic coextinction models: role of intrinsic dependence on animal-pollination [Dataset]"; DIGITAL.CSIC; http://dx.doi.org/10.20350/digitalCSIC/8507</p>			

IDENTIFICADORES

IDENTIFICADOR URI	dc.identifier	uri	Automático
<p>Handle del registro</p>			

IDENTIFICADOR ID INTERNO	dc.identifier		Opcional
<p>ID interno del conjunto de datos</p> <p>Es recomendable para conjuntos de datos con una estructura compleja y compuesta de distintos datasets.</p> <p>Ejemplo: SA4567-X</p>			



IDENTIFICADOR DOI	dc.identifier	doi	Automático
<p>Identifica el DOI del conjunto de datos</p> <p>DIGITAL.CSIC asigna DOIs a conjuntos de datos que se depositan en modalidad de acceso abierto o acceso embargado. En estos casos, este metadato se genera automáticamente varias horas después del depósito del conjunto de datos en el repositorio.</p> <p>Este metadato puede usarse también, y en ese caso no es automático, para incluir el DOI del dataset si ya lo tiene. Recomendado usar su forma enlazable.</p> <p>Ejemplos:</p> <p>http://dx.doi.org/10.20350/digitalCSIC/8507 http://dx.doi:10.5061/dryad.61s1n</p>			

TIPOLOGÍA

DESCRIPTOR	METADATO DUBLIN CORE	CUALIFICADOR	CARÁCTER
TIPO DE DOCUMENTO	dc.type		Obligatorio
<p>Un único valor posible: <i>Dataset</i></p>			

IDIOMA

IDIOMA	dc.language	iso	Obligatorio
<p>Idioma de los contenidos del conjunto de datos</p> <p>Según la ISO 639, con dos caracteres (EN, ES, IT..).</p> <p>Pueden seleccionarse más de un idioma pulsando la tecla Ctrl del ordenador.</p> <p>Elegir "No aplicable" cuando el dataset es, por ejemplo, una imagen o un video sin voz.</p>			

FILIACIÓN CSIC	dc.relation	csic	Obligatorio
<p>Producción o no del CSIC</p> <p>Para distinguir entre la producción institucional de la que es producción de los autores CSIC en períodos precedentes. La producción anterior al CSIC se deposita cuando así lo requieren los propios autores: cargas sistemáticas de CVs, si llegan a la biblioteca o recepción por SAD.</p> <p>Valores posibles: "Si" o "No".</p>			



ESTRUCTURA DE LOS DATOS

TABLA DE CONTENIDOS	dc.description	tableofcontents	Opcional
<p>Indica, por medio de un índice, los contenidos del conjunto de datos</p> <p>Recomendado para conjuntos de datos con una estructura compleja y compuestos de diversos datasets.</p> <p>Aparece en el formulario de carga en la última sección dedicada a la descripción de datasets.</p>			

DESCRIPCIÓN

DESCRIPTOR	METADATO DUBLIN CORE	CUALIFICADOR	CARÁCTER
PALABRAS CLAVE	dc.subject		Recomendado
<p>Palabras o frases que describen el tema o el contenido de los datos</p> <p>Suelen ser descriptores propuestos por los autores. Si aparecen en varios idiomas, primero se colocarán en inglés.</p> <p>Recomendado usar entradas procedentes de un vocabulario controlado. Si se usan entradas de Library of Congress Subject Headings (LCSH), Library of Congress Classification (LCC), Medical Subject Headings (MESH), Dewey Decimal Classification (DDC) entonces se recomienda usar, también, su metadato de clasificación correspondiente, disponible en el repositorio, con la URI de la entrada: dc.subject.lcsh, dc.subject.lcc, dc.subject.mesh, dc.subject.ddc.</p> <p>Deben evitarse los entrecorridos y las mayúsculas (excepto la letra inicial y en nombres propios).</p> <p>Metadato repetible tantas veces como palabras clave sean necesarias.</p> <p>Ejemplo:</p> <p>dc.subject: Cell Membrane dc.subject.mesh: https://www.ncbi.nlm.nih.gov/mesh/68002462</p>			

RESUMEN	dc.description	abstract	Recomendado
<p>Resumen de los objetivos y conclusiones de la investigación realizada</p> <p>Suele recoger la motivación del proyecto, el resumen del contenido del dataset y su estructura así como su metodología y proceso de validación.</p> <p>Hay varias cajas de texto para recogerlo en los posibles idiomas. Cada idioma precedido de [ES], [EN],...</p> <p>Si aparece en varios idiomas, primero debe ir el resumen en inglés.</p> <p>Cada caja de texto acepta hasta 4.000 caracteres.</p>			



INFORMACIÓN ADICIONAL	dc.description		Recomendado
<p>Aspectos complementarios sobre el conjunto de datos que no tengan cabida en otros campos:</p> <ul style="list-style-type: none"> Otras fechas: fecha de finalización del proyecto, fecha de lanzamiento, u otras fechas relacionadas con la vida útil de datos, por ejemplo, el ciclo de mantenimiento, actualización del programa Descripción de los contenidos de los ficheros Metodología detallada: ¿Cómo se generaron los datos, incluidos los equipos o el software utilizado, el protocolo experimental, etc? Acceso y reutilización: ¿Dónde y cómo los datos pueden ser accesibles por otros investigadores?, software necesario para leer los datos y licencia de uso (Creative Commons, Opendata License..). <p>Cada dato puede estar separado por .--</p>			

FINANCIACIÓN Y MANDATOS DE ACCESO ABIERTO

DESCRIPTOR	METADATO DUBLIN CORE	CUALIFICADOR	CARÁCTER
PATROCINADORES	dc.description	sponsorship	Recomendado
<p>Las organizaciones o agencias que financiaron la generación del dataset</p> <p>Sólo los que han contribuido económicamente a la generación y la gestión de los datos.</p> <p><i>We acknowledge financial support from the Spanish Science Ministry (MINECO) through projects TEC2011-29120-C05-04; and from European Research Council through Starting Grant NANOFORCELLS (ERC-StG-2011-278860).</i></p>			

AGENCIA FINANCIADORA	dc.contributor	funder	Recomendado
<p>Nombre normalizado de todas las agencias que financian el dataset</p> <p>Es obligatorio consultar el índice normalizado a la hora de incluir las agencias.</p> <p>Cada agencia va en un cuadro de texto.</p> <p>Agencias financiadoras: guía de normalización</p>			

NÚMERO DEL ACUERDO DE SUBVENCIÓN FP7	dc.relation		Opcional
<p>Número de la subvención FP7/Horizonte 2020 de la Comisión Europea, Ley de la Ciencia Española y la Comunidad de Madrid</p> <p>Obligatorio si el proyecto ha sido financiado por el FP7 u Horizonte 2020 de la Comisión Europea o el Consejo Europeo de Investigación (ERC) así como por la Ley de la Ciencia Española y la Comunidad de Madrid.</p> <p>Obligatorio para ser recolectado por el agregador europeo OpenAire (en el caso de la Unión Europea), Recolecta (Ley de la Ciencia Española) y e-ciencia (Comunidad de Madrid).</p>			



FP7/Horizonte2020. El formato debe ser **info:eu-repo/grantAgreement/EC/FP7/XXXXXX** o **info:eu-repo/grantAgreement/EC/H2020/XXXXXX**
 XXXXXX hace referencia al número del proyecto (grant agreement) y es siempre un número (ni acrónimos ni abreviaciones).

Ejemplo: info:eu-repo/grantAgreement/EC/FP7/226487

Ley de la Ciencia Española. El formato debe ser **MINECO/ICTI2013-2016/Referencia**

Referencia se refiere a la referencia del proyecto.

Ejemplo: MINECO/ICTI2013-2016/MAT2013-40489-P

Mandato OA de la CAM. Aquí escribiremos la referencia del proyecto: SYYYY/ABC-XXXX/Acrónimo.

FUENTES Y RELACIONES

DESCRIPTOR	METADATO DUBLIN CORE	CUALIFICADOR	CARÁCTER
FUENTES	dc.relation	isbasedon	Recomendado

Indica si el conjunto de datos se basa en conjuntos de datos previos, propios o ajenos. También puede indicarse otros estudios preliminares, de cualquier tipo, que se hayan utilizado como base para la generación del dataset

Aparece en el formulario de carga en la última sección dedicada a la descripción de datasets.

Ejemplos:

dc.relation.isbasedon: Spanish Meteorological Agency (AEMET); Ministry of Agriculture and Environment; Servei Meteorològic de Catalunya (METEOCAT); Navarra Government; SAIH Cantábrico; SAIH Duero; SAIH Ebro; SAIH Guadalquivir; SAIH Hidrosur; SAIH Júcar; SAIH Miño-Sil; SAIH Segura; SAIH Tajo

dc.relation.isbasedon: E. Lewis, D. Wallace, Program developed for CO2 system calculations. Oak Ridge, TN, USA: Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, 1–21. URL <http://cdiac.esd.ornl.gov/oceans/co2rprtnbk.html> (1998); C. Mehrbach, C. H. Culberson, J. E. Hawley, R. M. Pytkowicz, Measurement of the apparent dissociation constants of carbonic acid in seawater at atmospheric pressure. *Limnol. Oceanogr.* 18, 897–907 (1973); A. Dickson, F. Millero, A comparison of the equilibrium constants for the dissociation of carbonic acid in seawater media. *Deep Sea Res. A* 34, 1733–1743.30 (1987)

FUENTES	dc.relation	isreferencedby	Recomendado
---------	-------------	----------------	-------------

Referencia a los trabajos (por ejemplo, artículos) que mencionan el conjunto de datos

Podemos indicar, si se diera el caso, los trabajos que han mencionado (citan/referencian) el dataset que se está subiendo a Digital.CSIC. Es un metadato obligatorio si el dataset en cuestión es material suplementario de un artículo de revista.

Cada cita irá en un campo distinto (campo repetible).

Si estos trabajos se encuentran disponibles en Digital.CSIC se añadirá el handle correspondiente.

Por su importancia, este campo es visible en el registro resumido de Digital.CSIC (etiquetas).

¿Cómo se citan los trabajos?

- Si el trabajo tiene handle en Digital.CSIC: Apellido, Nombre; Apellido, Nombre. Título. DOI. Handle
- Si el trabajo no tiene handle pero si DOI: Apellido, Nombre; Apellido, Nombre. Título. DOI
- Si el trabajo no tiene DOI pero si handle: Apellido, Nombre; Apellido, Nombre. Título. Handle
- Si el trabajo no tiene ni handle ni DOI:

Apellido, Nombre; Apellido, Nombre. Título del trabajo. Citación (*)

(*) Para la citación ver campo 10 de la ["Guía para depositar trabajos en Digital.CSIC"](#)



ES PARTE DE	dc.relation	ispartof	Recomendado
<p>Indica si el conjunto de datos forma parte de otro, alojado o no en Digital.CSIC, o de otro trabajo de investigación</p> <p>Aparece en el formulario de carga en la última sección dedicada a la descripción de datasets.</p>			

ES VERSIÓN DE	dc.relation	isversionof	Recomendado
<p>Indica si el conjunto de datos es la versión de un conjunto de datos precedente</p> <p>Debe incluirse la URI del dataset previo o en su defecto, la citación bibliográfica de un dataset asociado precedente.</p> <p>Metadato solo disponible en modo de edición.</p>			

FORMATO

FORMATOS	dc.format	mimetype	Automático
<p>Tipo de formato</p>			

TAMAÑO	dc.format	extent	Automático
<p>Hace referencia a los bytes del conjunto de datos</p>			

SOFTWARE	dc.relation	requires	Recomendado
<p>Esta campo indica si es necesario un software específico para abrir, visualizar y/o reutilizar los datos</p> <p>Recomendado indicar dónde está accesible ese software, incluyendo su URL o si está depositado en Digital.CSIC, su handle.</p> <p>Aparece en el formulario de carga en la última sección dedicada a la descripción de datasets.</p>			

INFORMACIÓN ADICIONAL	dc.format		Recomendado
<p>Información sobre el formato/s en que se hace disponible el conjunto de datos</p> <p>Una buena práctica es ofrecer los datos en varios formatos, aquéllos que se usan más en una disciplina específica y después en un formato abierto.</p> <p>Metadato solo disponible en modo de edición.</p>			

DERECHOS DE USO

DESCRIPTOR	METADATO DUBLIN CORE	CUALIFICADOR	CARÁCTER
DERECHOS	dc.rights		Obligatorio
<p>Selección de la categoría pertinente: closedAccess, openAccess, embargoedAccess.</p> <p>Acceso abierto: tenemos el fichero del dataset. Acceso restringido: no tenemos el fichero del dataset. Embargado: tenemos el fichero del dataset pero está sujeto a un período de embargo hasta que se pueda hacer público. Para embargar un trabajo consulte la siguiente guía: Guía para embargar registros en Digital.CSIC</p> <p>Para añadir la fecha de embargo, consulte esta plantilla en el apartado "Fechas".</p>			

LICENCIA DE USO	dc.rights	license	Recomendado
<p>Indica la URL con más información sobre las condiciones de uso y reutilización</p> <p>Este campo está pensado para contenidos que han nacido para ser de acceso abierto (1) o contenidos que han sido creados con una licencia concreta (2) (por ejemplo, para literatura gris, material divulgativo o contenidos especiales sujetos a una licencia concreta).</p> <p>Se recomienda rellenar este dato para trabajos sujetos a licencias Creative Commons, Open Data Commons u otras.</p> <p>Ejemplos: http://creativecommons.org/licenses/by-nc-sa/4.0/ http://opendatacommons.org/licenses/dbcl/1.0/</p>			

REVISIÓN DE PARES

PEER-REVIEWED	dc.description	peerreviewed	Opcional
<p>Se consigna SI para los conjuntos de datos que hayan pasado un proceso de revisión/validación por una comunidad científica</p>			

IMPORTANTE: Inclusión de un readme file

- Proporciona información sobre el dataset para que sea correctamente interpretado por personas y máquinas.
- **Buenas prácticas:** un *readme* file por fichero de dataset, cuando sea posible, o uno muy desarrollado cuando el item se compone de varios datasets; denominar al readme file de tal manera que pueda ser asociado al dataset fácilmente; en formato plano p.e txt.
- **Buenas prácticas:** breve descripción del dataset; contacto del investigador principal y para preguntas; fecha de recogida de datos y de creación del dataset; información geográfica de los datos; metodología y enlace a publicaciones y otra documentación; unidades de medida, protocolos, abreviaciones, códigos, símbolos asociados a los datos; licencia de uso; citación recomendada

**IMPORTANTE:** denominación de los ficheros a adjuntar

Para una buena gestión de los recursos electrónicos y su accesibilidad es fundamental denominar adecuadamente los ficheros de los registros. A la hora de poner nombre a los ficheros se recomienda:

- Elegir nombres que sean suficientemente descriptivos del contenido, por ejemplo, usando parte de la referencia bibliográfica del trabajo en sí
- Evitar llamar a los ficheros por un número, ya que no es suficientemente descriptivo del contexto del registro
- Evitar el uso de caracteres como , . : ; / () \$ & | [] * < > " ¿
- Usar el guión bajo (_) mejor que espacios en blanco entre palabras
- No superar los 25 caracteres
- Para registros que tienen varias versiones, es aconsejable diferenciar unas versiones de otras indicando v01, v02 etc, en vez de update, nuevo etc. Una excepción resulta cuando denominamos a la versión definitiva, en cuyo caso puede llamarse FINAL. Para estos casos, se aconseja ser consistente en el nombre de los distintos ficheros para guardar una coherencia interna.