GUÍA PARA LA CREACIÓN DE DATASETS EN EL REPOSITORIO E-CIENCIADATOS

Índice

GLOSARIO....3
CARACTERÍSTICAS DEL REPOSITORIO E-CIENCIADATOS....3
PANTALLA PRINCIPAL DEL REPOSITORIO....4
CÓMO DEPOSITAR SUS DATOS DE INVESTIGACIÓN....5

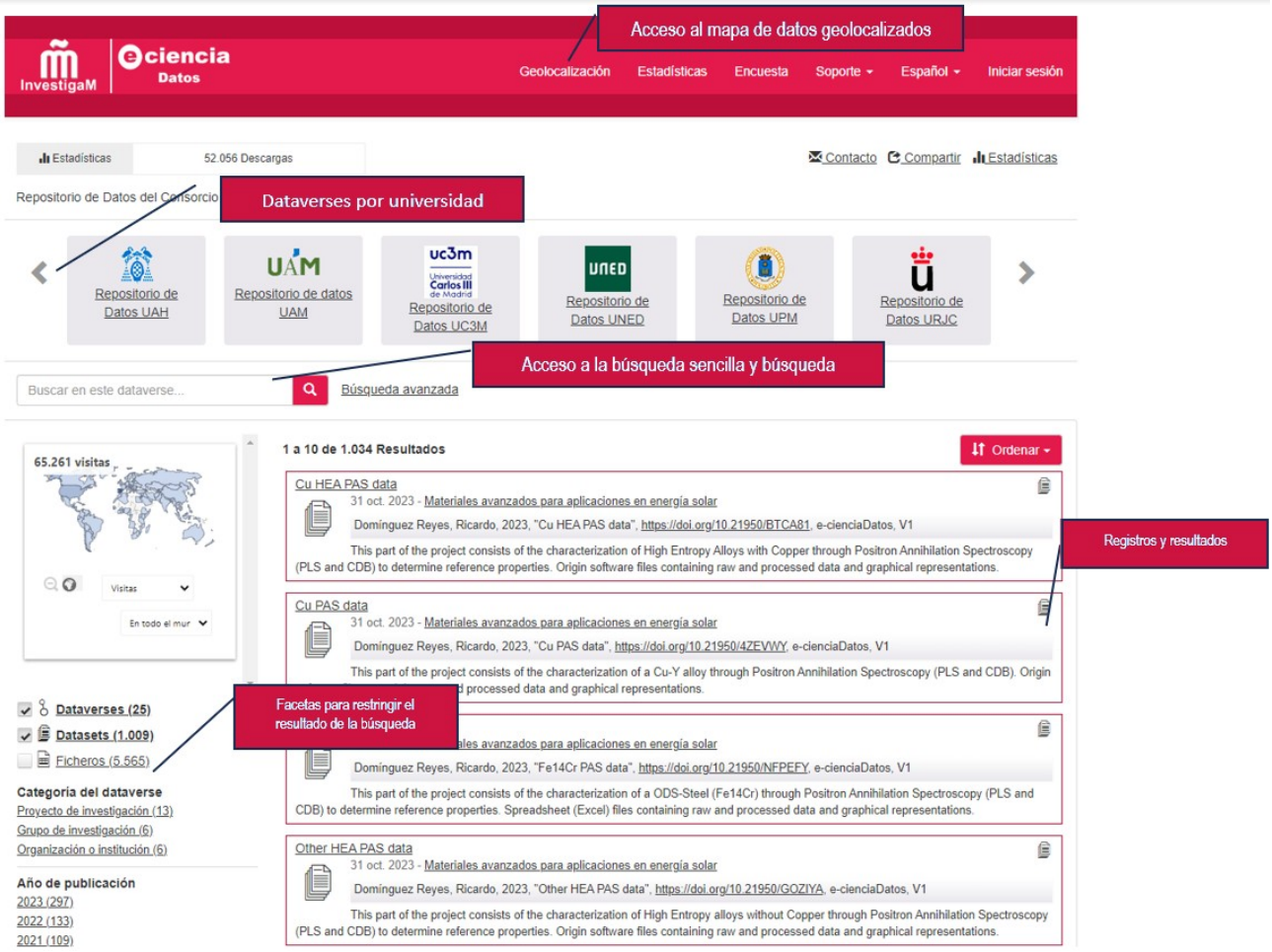
GLOSARIO

- e-cienciaDatos: Repositorio de datos multidisciplinar que alberga los conjuntos de datos científicos de los investigadores de las **universidades miembro del Consorcio Madroño**. ~~universidades públicas de la Comunidad de Madrid.~~.
- Dataset: Conjunto de datos de investigación.
- Comunidad Dataverse: Contenedor de varios datasets o conjuntos de datos para casos de proyectos y grupos de investigación que tengan varios conjuntos de datos o datasets.
- Metadato: Dato sobre el dataset, usado para su descripción y localización.
- DOI (*Digital Object Identifier*): Identificador único y persistente de un objeto digital (artículo electrónico, capítulo de libro electrónico, etc.).

CARACTERÍSTICAS DEL REPOSITORIO E-CIENCIADATOS

- Se estructura como un **sistema centralizado** constituido por distintas comunidades que agrupan los datasets o conjuntos de datos de la Universidad.
- El depósito de datos se hace mediante **archivo delegado**.
- Los datos están en **acceso abierto por defecto**, salvo razones legales, éticas o de transferencia de tecnología que así lo impidiesen.
- Se depositan los **datos finales de investigación** o datos que validen los resultados de la investigación.
- Cada dataset deberá ir acompañado de un **fichero de texto en formato txt (texto plano), denominado Readme.txt**, que contenga los metadatos necesarios para crear el dataset y que los datos sean comprensibles y reutilizables Ej: autoría, título, descripción, metodología, palabras clave, cobertura geográfica y temporal, software y licencia de uso, etc
- Se admite **cualquier formato de dato**, aunque se recomienda que se sigan **formatos estandarizados y lo más abiertos posible**.
- El repositorio **asigna un DOI** cuando se crea el dataset.
- Proporciona una **cita normalizada** para cada dataset para identificarlo y localizarlo.
- El repositorio e-cienciaDatos incluye **control de versiones**, lo que permite guardar el historial de versiones y ver el contenido de versiones anteriores de cada dataset.
- Hay que **leer y aceptar la licencia de depósito** que hace referencia a la cesión de los derechos de explotación de manera no exclusiva al repositorio, por lo que se podría también depositar los datos en otras plataformas (repositorios temáticos, webs, etc.).
- Se debe **seleccionar una licencia de uso** que indique las condiciones de uso, distribución y reutilización de los *dataset.* También se podrá aplicar licencias específicas para software Open Source Iniciative.
- El dataset **se creará como borrador (Draft), hasta que sea publicado por el administrador** de la comunidad dataverse de su universidad.
- e-cienciaDatos tiene el **certificado CoreTrustSeal** que lo acredita como repositorio de datos confiable (*trusted data repository*).

PANTALLA PRINCIPAL DEL REPOSITORIO



VERSIONES

Permite ver las diferentes versiones del dataset, en caso de que se realicen modificaciones en los metadatos, ficheros, etc. Para ver las diferencias entre una versión y otra, hay que seleccionar la versión del dataset y pulsar sobre Ver diferencias .

CÓMO DEPOSITAR SUS DATOS DE INVESTIGACIÓN

Puede depositar sus datos de investigación el personal investigador de las universidades miembro del Consorcio Madroño.

El flujo de trabajo del depósito es el siguiente:

1. El personal investigador comunica a su biblioteca la intención de publicar un dataset.
2. El responsable de su biblioteca administrador de la colección de la universidad se pone en contacto con el investigador, le informa sobre las buenas prácticas del repositorio, de las políticas de preservación y sobre datos sensibles además de informarle sobre el formato de datos recomendados por e-cienciaDatos. También le envía el enlace a la licencia de depósito y el plantilla Readme.txt. Este fichero tiene una versión en español y una versión en inglés en función de la audiencia a la que va dirigida el dataset y a la elección del investigador.
3. El personal investigador envía los ficheros del dataset a la biblioteca de su universidad junto con el plantilla Readme.txt (que incluye los metadatos) relleno y confirma que acepta la licencia de e-cienciaDatos. Si los datos no cumplen con los requisitos de e-cienciaDatos, el administrador responsable de su biblioteca se pondrá en contacto por correo electrónico para hacérselo saber de forma que el investigador realice las modificaciones oportunas.
4. El administrador de la universidad crea el dataset con los metadatos necesarios y el plantilla Readme.txt.
5. El responsable de la biblioteca revisa los datos y metadatos y si cree conveniente algún cambio, avisa al personal investigador. Además, le envía una url privada para que el investigador pueda validar los metadatos, ficheros y licencias del dataset como paso previo a la publicación. También se comunica al personal investigador cuál será el DOI del dataset.
6. Una vez el personal investigador y el responsable de la biblioteca están satisfechos con el resultado y se cumplen todos los requisitos para la publicación, se publica el dataset y se avisa al investigador.

Se pueden depositar conjuntos de datos de cualquier disciplina. Datos finales o datos completos, preparados para su distribución pública y su reutilización. Deberán contener la pertinente documentación anexa (Readme.txt).

¿Cuándo depositar sus datos de investigación?

Si están financiados por fondos públicos, cuando lo determine la entidad financiadora.

Los datos que validen resultados de investigación, lo antes posible.

Los otros datos susceptibles de ser compartidos que figuren en el Plan de Gestión de Datos.

Licencias de depósito y uso

Mediante la licencia de depósito se autoriza el depósito del dataset en e-cienciaDatos según los términos y condiciones indicados en la misma.

La licencia de uso es un instrumento legal, mediante el cual, el titular de los derechos de propiedad intelectual permite a terceras personas realizar determinados usos de los datos sin infringir dichos derechos. Hay disponibles diversas licencias libres, siendo las más

habituales las CC0 y CC-BY. Los datos estarán en acceso abierto.

¿Qué formatos de ficheros se admiten?

Formatos abiertos no propietarios para garantizar la accesibilidad y preservación y formatos propietarios que tienen un uso extendido entre las comunidades científicas que tienen interés en los datos.

Se puede subir todo tipo de ficheros en su formato original, aunque se recomienda que se usen **formatos estandarizados y lo más abiertos posible**.

Datos FAIR

Los datasets deben ser localizables, accesibles, interoperables y reutilizables, para lo que tienen metadatos descriptivos estandarizados, licencias de uso, documentación explicativa (Readme.txt) y DOI.

Valores añadidos

e-cienciaDatos se basa en el asesoramiento por parte de las bibliotecas, cumple con los requisitos de OpenAire y ofrece estadísticas y geolocalización. Permite descargar los metadatos en varios formatos compatibles con los principales gestores bibliográficos para citarlos correctamente.