

---

## Cargas retrospectivas de conjuntos de datos en DIGITAL.CSIC

Retrospective data upload to DIGITAL.CSIC

**Juan Román Molina**

[\(j.roman@bib.csic.es\)](mailto:j.roman@bib.csic.es)

Unidad de Recursos de Información Científica  
para la Investigación (URICI), CSIC, Madrid

---

Recibido: 7-5-2024; Revisado: 7-5-2024; Publicado: 24-5-2024

**Resumen:** DIGITAL.CSIC ha llevado a cabo la carga retrospectiva de conjuntos de datos de autoría CSIC de los años 2015 al 2022 tras una selección de fuentes gracias a la financiación recibida dentro de la convocatoria de ayudas para el fomento de la investigación científica de excelencia María de Guzmán 2020-2021.

**Palabras clave:** Conjuntos de datos; Datos de investigación; DIGITAL.CSIC; Carga retrospectiva; Principios FAIR

**Abstract:** DIGITAL.CSIC has carried out the retrospective loading of CSIC authoring datasets from 2015 to 2022 after a selection of sources thanks to the funding received within the call for grants for the promotion of scientific research of excellence María de Guzmán 2020-2021.

**Keywords:** Datasets; Research data; DIGITAL.CSIC; Retrospective data upload; FAIR Principles

**Como citar/Citation:** Román Molina, J. (2024). Cargas retrospectivas de conjuntos de datos en DIGITAL.CSIC. *Enredadera: revista de la Red de Bibliotecas y Archivos del CSIC*, (40), 59-62. <https://doi.org/10.20350/digitalCSIC/16286>

---

### Introducción

La comunidad científica internacional se encuentra inmersa en plena transición de paradigma sobre cómo comunicar los resultados de investigación. La Ciencia Abierta aboga por crear un sistema global de comunicación científica más abierto, transparente, colaborativo y sostenible persiguiendo un mayor impacto de los resultados de investigación con el fin de abordar los grandes retos de nuestra sociedad.

Este cambio de paradigma se caracteriza por el impulso del acceso abierto a publicaciones, la multiplicación de leyes y mandatos de acceso abierto de agencias financiadoras, universidades e instituciones científicas y la apertura y la explotación de los datos de investigación. Los repositorios de acceso abierto son un elemento clave en la materialización de este nuevo sistema.



El CSIC está comprometido con el acceso abierto a sus resultados de investigación desde la firma de la Declaración de Berlín en 2006. A lo largo de la última década este compromiso se ha reforzado paulatinamente, con la consolidación de su repositorio DIGITAL.CSIC y su participación en la red global de repositorios de investigación [COAR](#), con el apoyo a la iniciativa internacional [OA2020](#) a favor de nuevos modelos de negocio editoriales según los principios del acceso abierto, con la participación activa en la construcción de la Nube Europea de Acceso Abierto ([EOSC](#)) y con un papel destacado en las agendas de trabajo de [Science Europe](#) y del sistema nacional de investigación.

El [mandato institucional de acceso abierto del CSIC](#) dispone que a partir del 1 de abril de 2019 sean hechas públicas y de manera permanente a través de DIGITAL.CSIC las referencias bibliográficas de todas las publicaciones revisadas por pares desde el momento de la aceptación editorial para su publicación y que sus textos completos estén disponibles en acceso abierto en DIGITAL.CSIC tan pronto como sea posible. Igualmente, dispone que las referencias bibliográficas de los datasets asociados a artículos de revistas sean hechas públicas de manera permanente en DIGITAL.CSIC desde el momento de la aceptación para su publicación de los artículos asociados y que tales datasets se ofrezcan en acceso abierto en DIGITAL.CSIC siempre y cuando no se den legítimas razones de confidencialidad, propiedad intelectual y/o seguridad.

Además, todas las descripciones de estos resultados de investigación deben de ir en línea con los Principios FAIR.

### **Carga retrospectiva de conjuntos de datos en DIGITAL.CSIC (2015-2022): una línea de actividad en el recién terminado proyecto María de Guzmán del CSIC**

DIGITAL.CSIC, dentro de la convocatoria de ayudas para el fomento de la investigación científica de excelencia María de Guzmán 2020-2021 (punto 2 Ciencia en abierto e interoperabilidad en infraestructuras digitales institucionales, 2.1. Repositorios institucionales), recibió financiación para llevar a cabo una batería de desarrollos tecnológicos y acciones de apoyo documental y formativo en que se enmarca la estrategia para identificar conjuntos de datos de autoría CSIC para incorporarlos en el repositorio institucional.

DIGITAL.CSIC alberga colecciones de datos de investigación desde 2010, siendo la cuarta tipología más representada. Asigna DOIs a datasets desde 2016 y ofrece distintos servicios de apoyo para preparar planes de gestión de datos.

Sin embargo, DIGITAL.CSIC no reflejaba toda la producción de datos de investigación ya que estos se encontraban desperdigados por múltiples repositorios como Dryad y Zenodo. Esta carga retrospectiva mejora la representación en el repositorio de esta tipología documental con la intención de cumplir con las disposiciones del mandato CSIC de acceso abierto. Otro motivo,

de gran impacto para los gestores de repositorios, apunta a la dificultad de identificar las colecciones de datos de investigación de autoría institucional.

La Oficina Técnica de DIGITAL.CSIC ha aumentado manualmente, según unos criterios previamente establecidos, las colecciones de datasets presentes en el repositorio de forma retrospectiva desde 2015. Al mismo tiempo, las bibliotecas de la Red de Bibliotecas y Archivos del CSIC han estado subiendo conjuntos de datos al repositorio y los propios investigadores a través del Servicio de Archivo Delegado (SAD), encargando la carga de los datasets a las bibliotecas.

Los conjuntos de datos al estar disponibles en DIGITAL.CSIC aumentan su difusión, visibilidad y reutilización además de cumplir con el mandato institucional de acceso abierto.

Las fuentes seleccionadas para el descubrimiento fueron los repositorios [Zenodo](#) y [Dryad](#) y los agregadores [DataCite Commons](#), [OpenAire](#) y [Scholix](#).

Los trabajos de carga comenzaron por el agregador de datos [DataCite Commons](#) ya que albergaba bastante producción CSIC y [Zenodo](#), al ser este un repositorio utilizado por parte de la comunidad institucional para la difusión de resultados de proyectos enmarcados en iniciativas *multipaís*.

En estas plataformas se llevaron a cabo estrategias de búsqueda concretas según las características de cada una de ellas. Por lo general estas incluían el nombre de la institución, la tipología y los años (2015-2022). Por ejemplo, en DataCite Commons:

"Spanish National Research Council" en "Organizations", filtrando dataset en tipo de recurso, por años (2015-2022) e idioma inglés.

Y en Zenodo la ecuación de búsqueda fue:

CSIC OR "C.S.I.C." OR (cons\* AND sup\* AND invest\* AND cient\*) OR (spanish AND nat\* AND resear\* AND council), filtrando por dataset y registros "open". No era posible filtrar por año, pero nos ceñimos a 2015-2022.

El total de conjuntos de datos localizados de autoría CSIC, suma de las dos fuentes anteriores, ascendieron a poco más de 1.500.

Una vez localizados, los datasets fueron subidos al repositorio comprobando previamente que no lo estaban ya. Se usó la carga a través del propio repositorio DSpace utilizando la plantilla básica y/o la carga a través de los ficheros RIS (formato de archivo de etiquetas para el intercambio de datos bibliográficos mediante programas de citas) disponibles en algunas de las plataformas.

Para la descripción de los datasets se han seguido las normas de descripción de trabajos usadas en la Oficina Técnica de DIGITAL.CSIC, concretamente la [guía para depositar trabajos](#) y las plantilla de conjuntos de [datos](#) y [software](#) (por ejemplo <http://hdl.handle.net/10261/280578>).

Unos ejemplos de carga serían:

<http://hdl.handle.net/10261/283352>

<http://hdl.handle.net/10261/285520>

Los trabajos de carga comenzaron en septiembre de 2022 y aún continúan.

### **Datasets en la 4ª edición del mandato de acceso abierto del CSIC**

Con la misma finalidad de aumentar el número de datasets en el repositorio y para mejorar la presencia de datasets en DIGITAL.CSIC en la [4ª edición del monitor del Mandato CSIC de Acceso Abierto](#) se llevó a cabo la carga de conjuntos de datos presentes en artículos CSIC afectados por el mandato de acceso abierto. Se repasaron los trabajos revisados por pares pertenecientes a los institutos adheridos al Plan 100% Digital en 2022. Para este fin se elaboró un documento que explicaba todas las [casuísticas que se podían encontrar cuando el dataset estaba publicado junto al artículo](#). Estas cargas ascendieron a 1.632, llevando al año 2022 a la primera posición dentro de los años afectados por el mandato de acceso abierto del CSIC (desde 2019).