

Ibersid: XXVIII edición.

Implementación de los repositorios de datos de investigación en las universidades públicas españolas: estado de la cuestión

Francisco Javier MARTÍNEZ MÉNDEZ (UMU), Ana Alice BAPTISTA (UDM), Rosana LÓPEZ CARREÑO (UMU)

Ángel María DELGADO VÁZQUEZ (UPO)



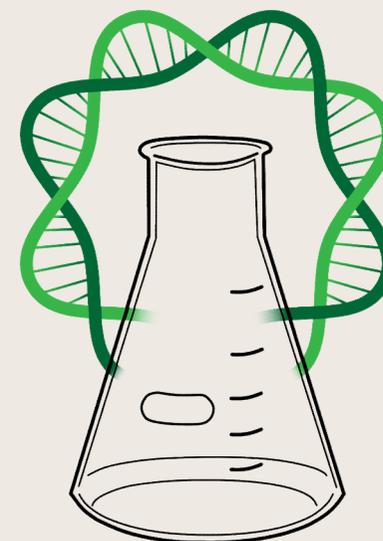
UNIVERSIDAD
DE MURCIA

Facultad de Comunicación
y Documentación

La **Ciencia Abierta** es uno de los pilares para el futuro de las civilizaciones.

Este movimiento aspira a hacer más abierto, accesible, eficiente, transparente y beneficioso el conocimiento científico (**UNESCO**).

En este contexto, los **conjuntos de datos de investigación**, siguiendo la metáfora de los **principios FAIR**, cobran especial importancia para mejorar la encontrabilidad, reproducibilidad y transparencia de la ciencia.



open science



UNIVERSIDAD
DE MURCIA

Facultad de Comunicación y
Documentación

La **Ciencia Abierta** facilita la replicación de los estudios, la verificación de las conclusiones presentadas en ellos y promueve una mayor transparencia de la actividad científica, algo de suma importancia porque la mayor parte de la investigación se financia con dinero público y ante ciertas prácticas poco edificantes que se vienen detectando.

Tiene en consideración la contribución de la **i+d+i** a la consecución de los **ODS** de la Agenda 2030 de **Naciones Unidas**.



Contexto de la investigación

La **Ciencia Abierta** representa un nuevo paradigma que incorpora una visión holística del proceso de generación de conocimiento, a partir del diseño de un proyecto y su desarrollo, hasta la comunicación, difusión y preservación de sus resultados (Abadal et al, 2023).

En la **Ciencia Abierta** viene cobrando impulso la idea de publicar, junto al artículo o informe derivado de la investigación, **el conjunto de datos** empleado en la misma con vistas a favorecer su reutilización por otros investigadores.

Esta gestión se halla dentro de las **principales líneas de trabajo** de las bibliotecas académicas (Federer & Qin, 2019), si bien no de manera estándar.

Acceder a los datos es imprescindible para reproducir estudios científicos, facilitar la cooperación, estimular el crecimiento económico y la innovación, aumentar la eficiencia de los recursos, mejorar la transparencia y **la confianza en la investigación científica** (OCDE 2021).

Los **conjuntos de datos** contienen información relevante y detallada que respalda los objetivos, análisis y resultados del estudio científico.

Los datos pueden ser primarios o secundarios, cuantitativos o cualitativos y estar disponibles en diferentes formatos y tamaños.

Es primordial que cumplan los **principios FAIR** ('findability – accessibility – interoperability – reuse') que establecen pautas para una correcta gestión de datos (Wilkinson et al., 2016).

Su gestión recae, en principio, en los investigadores, si bien está claro que necesitarán apoyo (Marín-Arraiza et al., 2019): una **oportunidad para reivindicar** el papel de los profesionales de la información.



Contexto de la investigación (II)

En España, la mayor parte de la investigación se lleva a cabo en universidades y centros públicos de *i+d+i* (Fundación CYD, 2023).

Son estas instituciones las que deben procurar implementar los repositorios de conjuntos de datos siguiendo las mejores prácticas posibles.

El **objetivo** de nuestro trabajo es analizar los repositorios de datos científicos de las universidades públicas españolas y establecer un **diagnóstico de su nivel de implantación**.

Directamente relacionadas con la gestión de repositorios de datos de investigación, como entidades de referencia, están también **REBIUN** y **FECYT**, juntos con dos consorcios autonómicos (Cataluña y Madrid).

En España, según los datos del **RUCT**, hay 86 universidades (50 públicas). Se excluye del estudio la **Menéndez Pelayo**. Quedan 34 universidades estudiadas individualmente y 14 que forman parte de los consorcios (8 y 6).

Dato	Propósito
Tipo	Si se trata de una universidad individual o de una que forma parte de un consorcio
URL consulta 'datasets'	Dirección del portal de consulta.
URL informativa	Dirección de la ayuda y documentación sobre los datos de investigación.
Modalidad	Si conforman un portal específicamente implementado para los datos de investigación o son una colección del repositorio institucional.
Software	La aplicación informática de soporte.
#datasets	Número de conjuntos depositados.
Archivo	Si es depósito directo por los investigadores o delegado en personal técnico.
Notas	Documentos de ayuda y guía al usuario.
Fecha/as de recogida de datos:	Primera revisión del 6 al 12 de marzo de 2023. Posteriormente, se revisaron datos hasta el día 25 de marzo de 2023.



Resultados

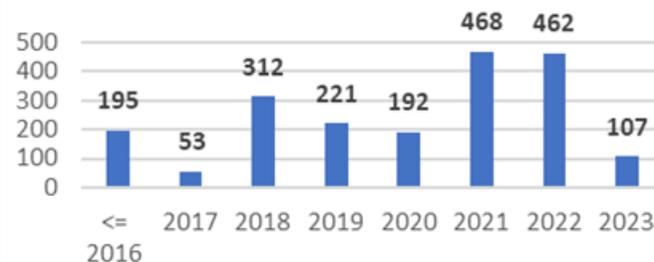
La revisión de los repositorios de conjuntos de datos de investigación muestra, en general, **un nivel de implementación muy diverso**, que oscila desde instalaciones mínimas (o inexistentes) hasta la implementaciones de repositorios colectivos muy completas.

Nombre	#univ	Composición	#datasets
CORA-RDR	8	Autónoma de Barcelona; Barcelona; Girona; Lleida; Oberta de Catalunya; Politècnica de Catalunya; Pompeu Fabra; Rovira i Virgili e Illes Balears	467
e-Ciencia-Datos	6	Alcalá de Henares; Autónoma de Madrid; Carlos III, UNED; Politècnica de Madrid y Rey Juan Carlos	744
Resto de universidades	25	A Coruña; Alicante; Almería; Burgos, Cádiz, Cantabria, Complutense, Córdoba, Granada, Huelva, Jaén, Jaime I, León, Málaga, Murcia, Oviedo, Pablo de Olavide, País Vasco, Politècnica de Cartagena, Politècnica de Valencia, Salamanca, Sevilla, Valencia, Valladolid y Zaragoza.	744

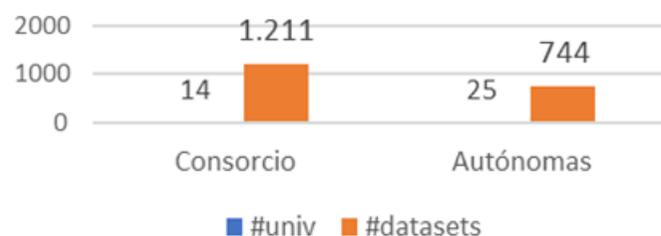


Resultados (II)

datasets por año de publicación

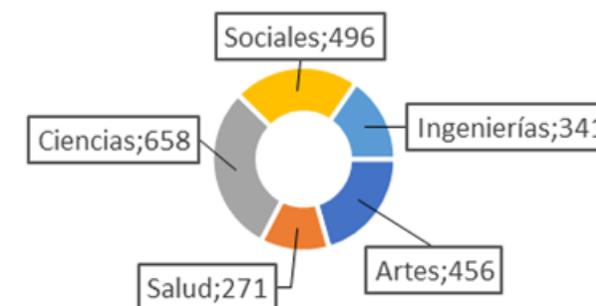


datasets por tipo de organización



Universidad	#datasets
Alcalá de Henares	375
Carlos III	259
Zaragoza	222
Granada	112
Barcelona	105
Politécnica de Catalunya	75
Lleida	59
Pompeu Fabra	44
Politécnica de Valencia	43
Politécnica de Madrid	41

datasets por materia

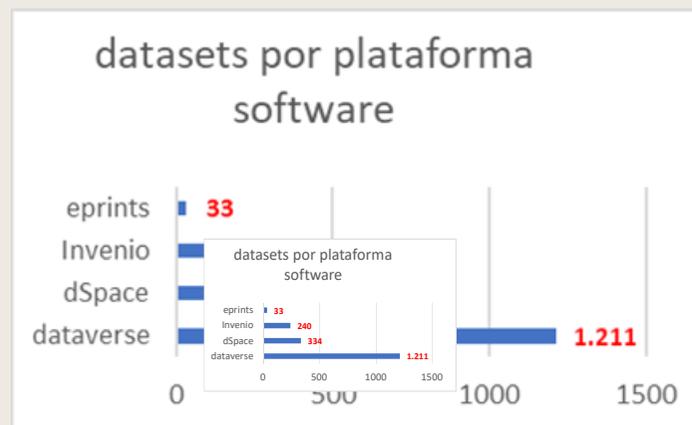


Resultados (III)

Registro de los conjuntos de datos

No ha sido sencillo encontrar esta información. En **los consorcios** el archivo es delegado (es uno de los servicios ofrecidos). En las universidades que gestionan de forma individual el repositorio, **12** optan por **autoarchivo**, **1** por **archivo delegado** y 3 por una combinación.

El predominio del **autoarchivo** es fruto, sin duda alguna, de la costumbre heredada del depósito en los repositorios institucionales.



Documentación y ayuda

36 de las **39** universidades con repositorio de conjuntos de datos **orientan y ayudan de forma suficiente** a su comunidad. Es previsible que las **3** universidades restantes solucionen este problema pronto.

Los **temas principales** son la gestión de los datos de investigación y cómo se planifica; información sobre aplicaciones para elaborarlos, orientación sobre los posibles lugares de depósito y sobre la importancia del uso de licencias tipo **Creative Commons**.



Resultados (IV)

Langer et al. (2019) analizaron varias plataformas de repositorios. Establecieron 3 categorías: genéricas para la gestión básica de los datos; aplicaciones como desarrollo de estas plataformas y otras herramientas de gestión. En las universidades españolas son todas **básicas**.

También investigaron sus capacidades de **gestión de datos sostenible** aplicando 15 criterios basados en los principios **FAIR**.



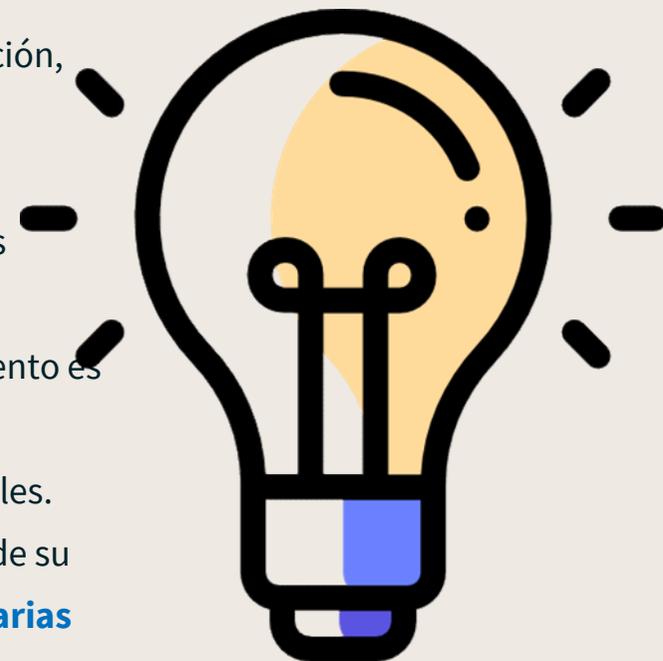
Subgrupo FAIR	Significado
F: Localizable	C1 ¿Se puede acceder a un determinado conjunto de datos de investigación en una versión actual a través de un PID/DOI único?
	C2 ¿Está la información de los datos de investigación depositados indexada en catálogos de datos, registros y motores de búsqueda?
	C3 ¿Existe una interfaz de búsqueda con posibilidades de filtrado para datos enlazados estructurados?

	F	A	I	R
Dataverse	C1+ C2+ C3o	C4+ C5+ C6+ C7o	C8+ C9+ C11o	C12+ C13+ C14o
DSpace	C1+ C2+ C3o	C4o C5o C6+ C7+	C9+	C12o C13+ C14o
eprints	C1+ C2+ C3o	C4+ C6+ C7o		C12+ C13+ C14o
Invenio	C1 + C2 + C3 o	C4+ C6+	C9o C11o	C13o C15+



Conclusiones

1. Estamos **casi en el inicio** de la implementación de los repositorios, salvo alguna excepción, excluyendo a los consorcios autonómicos (podría ser un ejemplo a seguir).
2. Este estado embrionario es una **oportunidad para unificar criterios** que eviten multiplicidad de esfuerzos similares y dispersión de los desarrollos e implementaciones (como pasó con repositorios generalistas).
3. Se está dando un gran impulso a la publicación de estos conjuntos de datos, el crecimiento es continuo.
4. El **autoarchivo** se impone. Debería **ser obligatoria la revisión** por parte del profesionales.
5. Una gestión de los conjuntos de datos, junto a una campaña de sensibilización a favor de su uso entre la comunidad investigadora, **reforzará el papel de las bibliotecas universitarias** en el apoyo a la investigación.
6. Hay que estudiar la (“**no**”) sostenibilidad de la publicación de conjuntos de datos en varios repositorios a la vez.





Muchas gracias por su atención