

# FAIR x FAIR. Una aproximación pragmática a la gestión de datos de investigación en modo FAIR

## FAIR x FAIR. A pragmatic approach to research data management in FAIR mode

Mireia Alcalá; Lluís Anglada

Alcalá, Mireia; Anglada, Lluís (2019). "FAIR x FAIR. Una aproximación pragmática a la gestión de datos de investigación en modo FAIR". *Anuario ThinkEPI*, v. 13, e13e05.

<https://doi.org/10.3145/thinkepi.2019.e13e05>

Publicado en *IweTel* el 6 de septiembre de 2019

### Mireia Alcalá

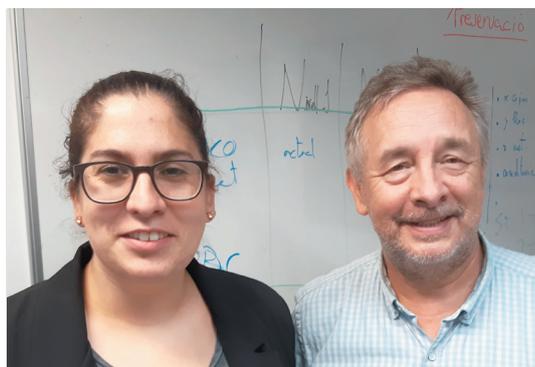
<https://orcid.org/0000-0003-2684-2110>

Consorci de Serveis Universitaris de Catalunya (CSUC)  
Gran Capità, 2 (Edifici Nexus). 08034 Barcelona  
[mireia.alcala@csuc.cat](mailto:mireia.alcala@csuc.cat)

### Lluís Anglada

<https://orcid.org/0000-0002-6384-4927>

Consorci de Serveis Universitaris de Catalunya (CSUC)  
Gran Capità, 2 (Edifici Nexus). 08034 Barcelona  
[langlada@gmail.com](mailto:langlada@gmail.com)



**Resumen:** En los últimos años la gestión de datos de investigación ha ocupado un lugar destacado en la agenda de la comunidad científica y, concretamente, de la *Comisión Europea*. En 2017 el *Consorci de Serveis Universitaris de Catalunya (CSUC)* creó un área específica para acometer de forma colectiva los retos derivados de la ciencia abierta. Se están realizando diferentes actividades sobre gestión de datos (formación, apoyo a la creación de planes de gestión de datos y políticas para repositorios de datos) de forma conjunta. Recientemente una comisión formada por los vicerrectores de investigación pidió un informe sobre cómo debería ser la infraestructura que la *Comisión Europea* pide para publicar los datos de investigación en forma

FAIR (*Findable, Accessible, Interoperable, Reusable*). El informe se hizo recogiendo la opinión de diversos expertos y se terminó en marzo de 2019. Es una aproximación pragmática a los requisitos que debería tener un repositorio de datos para cumplir con los principios FAIR. Paralelamente a los requisitos propiamente dichos, se determinan unos condicionantes de contexto y se establecen buenas prácticas a seguir. El informe finaliza con tres recomendaciones finales: crear un repositorio de datos, desarrollar buenas prácticas en la gestión de datos, e impartir formación sobre ciencia abierta y gestión de datos de investigación.

**Palabras clave:** Datos de investigación; Repositorios de datos; Principios FAIR; CSUC; Cataluña; Servicios bibliotecarios de apoyo a la investigación; Ciencia abierta; Informe.

**Abstract:** In the last few years research data management has been at the top of the scientific community agenda and, specifically, of the *European Commission (EC)*. In 2017, *Consortium of University Services of Catalonia (CSUC)* created an area aimed at collaborating with universities to collectively undertake the challenges of Open Science. Different activities are being carried out together related to data management (training, advise to create data management plans and policies for data repositories). Recently, a commission by the vice-rectors for research requested a report on how the infrastructure recommended by the EC should look like to publish research data fulfilling the FAIR principles (*Findable, Accessible, Interoperable, Reusable*). The report, completed last March, includes the opinion of various experts. It is a pragmatic approach to determine the requirements that a data repository should have to go FAIR. Alongside the requirements themselves, the document determines contextual conditions and establishes a set of good practices. The report concludes with three final recommendations: to create a data repository, to develop good practices in data management, and enable training in Open Science and research data management.

**Keywords:** Research data; Data repositories; FAIR principles; *Consortium of University Services of Catalonia*; *CSUC*; Library support services for research; Open Science; Report.

## 1. La publicación en abierto de los datos de la investigación

La ciencia siempre ha usado datos como materia prima para la investigación y el avance del conocimiento. Pero hasta hace poco, éstos eran difíciles de difundir, compartir y reutilizar. Actualmente la capacidad de procesar, conservar y comunicar de los ordenadores y las redes de comunicación ha cambiado profundamente el sistema de comunicación científica y lo ha hecho más eficiente, transparente y colaborativo. Así lo señala, por ejemplo, la *CRUE* cuando afirma que

“la actividad investigadora actual es intensiva en la generación, consumo y explotación de datos y su preservación y reutilización son clave en la investigación del siglo XXI” (*CRUE*, 2019).

Por este motivo se considera que los datos son el ‘nuevo oro’ del siglo XXI, se les asigna un alto valor estratégico y en estos últimos años están en el centro de la atención de la comunidad científica y de los órganos gestores de la investigación. La *Comisión Europea (CE)* con el programa marco *Horizonte 2020* y su plan piloto *Open Research Data Pilot* ha dado un gran impulso al movimiento conocido como “gestión de datos de investigación” (*research data management*, RDM<sup>1</sup>). Este es un concepto paraguas que engloba diferentes actividades relacionadas con la creación, organización, estructuración, almacenamiento, preservación y compartición de los datos.

La *CE* establece como requisitos para los proyectos que financia que tengan un plan de gestión de datos (*data management plan*, DMP) y que los datos se depositen en abierto siguiendo los principios FAIR (*Findable, Accesible, Interoperable y Reusable*). Estos principios significan que los datos deben ser encontrados (la F de FAIR) a través de la asignación de un identificador único y persistente y una correcta descripción mediante los metadatos. Una vez encontrados, los datos deben ser accesibles (la A) con o sin restricciones mediante procesos de autenticación y/o autorización y protocolos de comunicación abiertos. Pero para que los datos no queden en silos aislados deben interoperar (la I) no sólo con otros datos sino también con otros repositorios y aplicaciones. Finalmente, el gran valor de los datos es que puedan ser reutilizados (la R) para así acelerar los descubrimientos científicos y ahorrar costes en su confección.

---

**“La CE establece como requisitos para los proyectos que financia que tengan un plan de gestión de datos (*data management plan*, DMP) y que los datos se depositen en abierto siguiendo los principios FAIR (*Findable, Accesible, Interoperable y Reusable*)”**

---

## 2. El servicio de apoyo a la gestión de datos de investigación

Como todas las bibliotecas, las de las universidades de Cataluña han llevado a cabo servicios y actividades de apoyo a la investigación. Quizá la más importante fue la contratación conjunta de revistas a través del *Consorcio de Bibliotecas Universitarias de Catalunya (CBUC)*<sup>2</sup>. La importancia creciente de apoyar unos métodos científicos en pleno proceso de cambio llevó a que en 2013 se aprobara una línea específica de trabajo de apoyo a la investigación y que a principios del 2017 se creara dentro del consorcio un área específica dedicada a promover la ciencia abierta.

Esta área tiene por objetivo acometer de forma colectiva los retos derivados de la ciencia abierta y procurar que su adopción se realice con el menor esfuerzo por parte de las universidades. Está compuesta por dos comisiones:

- una más política formada por los vicerrectores de investigación;
- otra técnica con miembros de las bibliotecas y los servicios de apoyo a la investigación de las universidades miembros del *CSUC*.

Con respecto a la gestión de datos, en primera instancia los trabajos se centraron en alinear las distintas instituciones realizando actividades de formación a diferentes niveles: invitando a diferentes expertos como docentes de cursos para el personal que apoya la investigación. A su vez se han realizado encuestas a investigadores con la finalidad de conocer sus prácticas y necesidades con respecto a los datos de investigación.

A partir de aquí los trabajos se orientaron a dar respuesta a los dos requisitos que pedía la *CE*: el DMP y la publicación en abierto de los datos de investigación.

Respecto a los DMPs, se elaboró un instrumento que permite crear, compartir y exportar DMPs siguiendo los requisitos de la CE. Este instrumento (eiNa DMP) usa el software elaborado por el *Digital Curation Centre (DCC)* y *University of California Curation Centre (UC3)* y contiene descripciones y ejemplos reales de lo que debe incluir cada apartado del plan. La guía se mantiene de forma colectiva entre las diferentes universidades miembros del CSUC, pero puede visualizarse bajo diferentes formas de manera que puede adaptarse a una institución o grupos de usuarios concretos.

Con respecto a cómo hacer públicos los datos, se ofrecieron dos soluciones

- la primera era hacerlo en algún repositorio de los existentes; para ello se elaboró una relación de opciones disponibles con sus principales características (CSUC, 2017);
- la segunda ha sido adoptada por algunas universidades y ha consistido en la adaptación de sus repositorios institucionales para depositar ciertos tipos de datos a pesar de que se es consciente de que la solución no cumple con todos los principios FAIR (por falta de identificadores, prestaciones de preservación o capacidad alta de almacenamiento, por ejemplo).

Complementariamente, la comisión de vicerrectores aprobó documentos (CSUC, 2017) que definen los diferentes apartados que debe tener una política de acceso abierto a los datos de investigación con el objetivo de facilitar que las universidades u otras instituciones confeccionen sus propias políticas.

Después de dos años de ofrecer estos servicios de apoyo a la gestión de datos de investigación se evaluaron los resultados y se constataron dos carencias importantes:

- el poco uso de los servicios ofrecidos (cosa que puede ser debida más a la novedad del tema y al hecho de que la publicación de los datos no es obligatorio en la mayoría de casos);
- no se dispone de la infraestructura que la CE pide para publicar los datos de investigación en forma FAIR.

### **3. ¿Qué requisitos razonables debería tener un repositorio de datos?**

La comisión de vicerrectores analizó los resultados de estos dos años de trabajos y sus debilidades. Mientras que paliar la primera requería tiempo y paciencia, solucionar la segunda exigía claramente tener un repositorio dónde publicar los datos tal como ya lo están haciendo o prevén hacerlo algunos consorcios o países<sup>3</sup>. Pero ¿qué requisitos razonables debería tener un repositorio de datos y qué coste tendría hacerlo? La comisión encargó la elaboración de un informe que determinara los requisitos funcionales razonables que debe tener un repositorio de datos para que cumpla con los requisitos FAIR.

Para poder definir y estudiar los requisitos se estudiaron los principales documentos que se han publicado recientemente sobre este tema pero se comprobó que éstos eran más teóricos que prácticos, por lo que se creó un grupo de trabajo específico formado por miembros de los diferentes *stakeholders* de la universidad (biblioteca, oficinas de investigación, servicios TIC). Además, se entrevistó a 32 expertos nacionales e internacionales consultándoles las características que debería tener un repositorio de este tipo desde su punto de vista.

La intención inicial de las entrevistas con expertos era poder hacer una lista de requisitos, pero, desde las primeras reuniones se constató que adicionalmente a los requisitos técnicos, y a menudo en un orden de importancia mayor, los expertos destacaban dos elementos a tener en cuenta:

- determinar los condicionantes de contexto del repositorio;
- desarrollar unas buenas prácticas asociadas a la definición de conjuntos de datos FAIR.

### **4. Contexto y requisitos**

Los expertos señalaron reiteradamente que la gestión de datos de investigación y su publicación en forma abierta es un tema de largo recorrido, pero todavía incipiente. Esto conlleva que los requisitos que ahora se fijan haya que ampliarlos o modificarlos en los próximos años con toda probabilidad, pero no debería ser motivo para no crear de forma inmediata una infraestructura de datos de investigación ya que uno de los objetivos de tenerla es generar experiencia y buenas prácticas para su gestión.

En este sentido y para enmarcar el repositorio, el informe considera que debería hacerse ahora (para así poder desarrollar conocimiento práctico sobre la gestión de datos de investigación) y aquí (para facilitar el cumplimiento de las regulaciones legales). A su vez considera que debe distinguirse entre la publicación de datos permanentes o finales y la gestión de ficheros de datos provisionales y ceñirse a datos de disciplinas que no tienen repositorios de datos consolidados, ya que, para los que tienen, la mejor opción es publicarlos allí.

La parte principal del informe se centra en describir los requisitos funcionales que se consideran 'razonables' (es decir: factibles y realizables) para garantizar que la infraestructura creada cumpla los requisitos FAIR de la CE.

Los primeros requisitos los relacionamos con la asignación y uso de identificadores persistentes (DOIs para los datos y orcid para investigadores). También, y en relación con la capacidad de almacenamiento, el

repositorio debe admitir hasta 10 GB por defecto y debe tener un espacio elástico para que pueda crecer.

Por lo que se refiere a la preservación indicamos que el archivo debe estar disponible durante 10 años y tener dos copias geográficamente distribuidas como mínimo. Además, para garantizar la integridad de los datos, será obligatorio verificar periódicamente el *checksum*; todo ello, siguiendo el modelo de preservación OAIS.

Uno de los puntos clave es la interoperabilidad con otros sistemas, ya que el repositorio debe comunicar con otros (ya sean repositorios para datos o los institucionales) así como comunicar con herramientas de almacenamiento en la nube para permitir una carga amigable de los datos. Todas estas comunicaciones deben hacerse mediante protocolos de comunicación estándar, así como usando estándares de metadatos.

Finalmente -pero no menos importantes-, se detallaron una serie de requisitos relacionados con la gestión de características especiales como por ejemplo: gestionar diferentes esquemas de metadatos, los diferentes tipos de acceso, de ingesta, ofrecer la citación recomendada, entre muchos otros.

## 6. Buenas prácticas y recomendaciones finales

Está claro que la infraestructura es importante, pero para los datos de investigación, el acceso presente y su reutilización futura no dependen solamente de las prestaciones del repositorio, sino de características formales asociadas a los conjuntos de datos y de los procesos asociados a su confección. Debido al estadio aún incipiente de la gestión de datos, estas 'buenas prácticas' no están del todo establecidas, cosa que no las hace menos importantes.

El informe destaca su importancia y distingue como buenas prácticas las siguientes:

- hacer curación de datos para que sean comprensibles y reutilizables por otros usuarios distintos a los han generado;
- crear protocolos y criterios de expurgo y selección de los conjuntos de datos;
- fomentar el uso de formatos abiertos o no propietarios y, en todo lo que sea posible, migrar los formatos propietarios a formatos abiertos;
- usar estándares, protocolos y vocabularios controlados ampliamente aceptados.

Teniendo en cuenta todas las consideraciones anteriores, el informe resultante -publicado con el título de *FAIR x FAIR. Requisitos factibles, alcanzables e implementables para un repositorio de datos de investigación FAIR* (CSUC, 2019)- hace las recomendaciones finales siguientes:

- Crear un repositorio en el marco de las universidades catalanas de manera inmediata con software libre, donde se puedan publicar los datos de investigación de forma FAIR y que permita desarrollar experiencia y buenas prácticas en la gestión de datos de investigación.  
Promover y facilitar, desde las diferentes unidades implicadas, la publicación de los datos de investigación en abierto.
- Hacer formación sobre ciencia abierta y gestión de datos de investigación, distinguiendo entre usuarios avanzados, investigadores jóvenes y personal de apoyo de las universidades.

## 7. Notas

1. En algunos casos, quizá especialmente en EUA se usa el termino *research information management* (RIM).
2. En 2013, el CBUC se fusionó con el Centro de Servicios Científicos y Académicos de Catalunya (CESCA) para formar el Consorcio de Servicios Universitarios de Catalunya (CSUC).
3. En España, la única propuesta es *ecienciaDatos* del Consorcio Madroño. Por otro lado, los países europeos consultados con características similares a Catalunya han sido: Bélgica flamenca, Finlandia, Países Bajos, Portugal y Suecia

## 8. Bibliografía

CRUE (2019). *Compromisos de las universidades ante la Open Science*.  
[http://www.crue.org/2019.02.20-CompromisosCRUE\\_OPENSCIENCE.pdf](http://www.crue.org/2019.02.20-CompromisosCRUE_OPENSCIENCE.pdf)

CSUC (2017). *Proposta per establir una política d'accés obert a les dades de recerca a les Universitats de Catalunya*.  
[http://csuc.cat/sites/default/files/docs/1603politiques\\_rdm-v8.pdf](http://csuc.cat/sites/default/files/docs/1603politiques_rdm-v8.pdf)

CSUC (2017). *Recomanacions per seleccionar un repositori per al dipòsit de dades de recerca: Versió 3, Maig 2017*.  
<http://hdl.handle.net/2072/284974>

CSUC (2019). *Data management plans: Version 3, June 2019*.  
<http://hdl.handle.net/2072/357988>

CSUC (2019). *FAIR x FAIR. Requisitos factibles, alcanzables e implementables para un repositorio de datos de investigación FAIR*.  
<http://hdl.handle.net/2072/356460>