

# ¿Qué es la gestión de datos y cómo la relaciono con mi proyecto de investigación?

MESA DE DEBATE

28 DE SEPTIEMBRE DE 2022

## **REBECA MARÍN DEL CAMPO**

Universidad Carlos III de Madrid (España)

## **EVA ORTIZ**

Universidad Carlos III de Madrid (España)

### **RESUMEN EXTENDIDO**

Las expositoras hacen un recorrido en torno a los datos de investigación, los tipos existentes y sus conjuntos. Para esto proponen ejemplos como es el Proyecto Genoma Humano, en el cual las bases de datos se hallan a disposición de los científicos en el mundo, sin restricciones en cuanto a su uso o redistribución.

Basándose en la importancia de los datos en abierto y la utilidad de compartirlos, refieren la recomendación de la UNESCO (2021), lo que implica un cambio de paradigma en la gestión de los datos de investigación. Y mencionan que los datos en abierto verifican los resultados de investigación, permiten su reutilización; aumentan la credibilidad de la ciencia, pues hace patente la inversión pública en investigación; mejoran de la transparencia; aumentan la visibilidad y el impacto de la investigación; favorecen la generación de nuevos contactos entre colegas, impulsando la colaboración y finalmente contribuyen en la preservación.

Otro tema al que hacen referencia es al acrónimo FAIR, que en español se traduce como: “Encontrable”, “Accesible”, “Interoperable”, “Reutilizable”, establecidos por la iniciativa Force 11 para determinar cómo deben ser los datos al compartirse.

Posteriormente, se hace mención al tema de las políticas, las cuales resultan fundamentales porque crean un marco de actuación y además potencian el desarrollo y la implantación de la gestión de los datos. Además, las expositoras mencionan algunas iniciativas en el ámbito de la Unión Europea, como Horizonte Europa, la European Open Science Cloud, la Ley de España 17/2022 de la Ciencia, la Tecnología y la Innovación, que en su artículo 37 que hace referencia a la Ciencia Abierta. Y a nivel latinoamericano, las Políticas para la Ciencia Abierta y los Datos Científicos en América Latina y legislaciones nacionales de Perú, Argentina, México y proyectos en Chile y Brasil.

Tal información se complementa con el ciclo de vida de los datos de investigación, que corresponde con la gestión de datos de investigación (GDI) y, a su vez, implica una serie de aspectos tales como la propiedad y los tipos de datos, el almacenamiento, los formatos, las normas disciplinarias de metadatos, la estructura y convenciones, y la documentación. También se evidencian aspectos éticos y datos personales, para el uso adecuado de los datos «sensibles», y las buenas prácticas como la anonimización, que supone el procesamiento de los datos personales de forma que se prevenga de manera irreversible la identificación. Puede ser por razones éticas, como proteger la identidad.

Considerando los aspectos anteriores se puede proceder a la preparación de un Plan de Gestión de Datos (DMP/PGD), que es un documento formal que debe presentarse al inicio de la investigación y modificarse según se avanza con el proyecto, por lo tanto, es un documento vivo. Existen varias plataformas de apoyo para cumplimentarlo, todas basadas en un cuestionario donde explicar la gestión durante el ciclo de vida de los datos, desde su creación hasta la finalización del proyecto de investigación.

Algunos ejemplos de plantillas normalizadas para diferentes financiados (son de uso libre):

- Horizonte Europa: [Plantilla para DMP](#)
- PARTHENOS: [Plantilla para datasets de arqueología](#)
- Digital Curation Centre: [Checklist for a Data Management Plan](#)
- [Madroño PGD Online](#)
- [Argos](#)

Existen también buenas prácticas para el tratamiento de ficheros a depositar, como son: nombres, formatos, protección de datos personales, etc. para que estos sean significativos/descriptivos y consistentes. La recomendación es aplicarlas tanto para nombres, como para los distintos formatos.

El siguiente aspecto es el de las licencias de uso, que corresponden a la forma de comunicar qué permisos disponen los potenciales usuarios, las cuales están determinadas por las personas creadoras, productoras o distribuidoras de los datos. Para identificar cuáles son los tipos se puede acudir a Creative Commons, Open Data Commons, y para licenciar software a MIT License, Apache License, y GNU General Public License.

En cuanto al tema de los repositorios, se considera para la difusión de datos finales que sean «tan abiertos como sea posible, tan cerrados como sea necesario» y visualizar la importancia de los repositorios en la preservación.

Finalmente, se presentan ejemplos concretos de la experiencia en España, como el Consorcio Medroño, en el cual se han elaborado guías de cómo depositar los datos para el personal investigador de las universidades que forman parte del Consorcio, o el Grupo Técnico Iberoamericano de Dataverse conformado por universidades y consorcios. Y otro ejemplo en la Universidad Carlos III que también pone a disposición una serie de indicaciones para el depósito de datos en e-ciencia de datos.

[Ver conferencia en YouTube](#)