

Publicado en *blok de bid* (<http://www.ub.edu/blokdebid>)

[Inicio](#) > Alud de datos

## Alud de datos

Enviado por [Isabel Bernal](#) [1] el Mié, 08/01/2014 - 11:41. [Blok de BiD - esp](#) [2] [datos de investigación](#) [3] [Isabel Bernal](#) [4] [políticas](#) [5]

**Isabel Bernal**

**Unidad de Recursos de Información Científica para la Investigación  
CSIC**

---

Wilma Mossink, Magchiel Bijsterbosch, and Joeri Nortier. *European Landscape Study of Research Data Management*. Utrecht: SURF, 2013.

[http://www.sim4rdm.eu/sites/default/files/uploads/documents/SIM4RDM%20landscape%20report%20vs1%204\\_14.08.13.pdf](http://www.sim4rdm.eu/sites/default/files/uploads/documents/SIM4RDM%20landscape%20report%20vs1%204_14.08.13.pdf) [6]



SIM4RDM (*Support Infrastructure Models for Research Data Management*, <http://www.sim4rdm.eu/> [7]) es un proyecto europeo iniciado en octubre de 2011 y financiado por el 7 Programa Marco de la Comisión Europea que se encuadra en el creciente número de iniciativas que, ante el alud de datos científicos, promueven buenas prácticas y estándares en su gestión para garantizar el progreso científico y económico mediante su acceso, reutilización y preservación digital. En concreto, SIM4RDM se marca como objetivo facilitar a la comunidad científica el uso efectivo de infraestructuras emergentes para datos de investigación y parte de la consideración previa del relativo desconocimiento por los investigadores de metodologías, instrumentos y capacidades necesarias. Para ello, el consorcio SIM4RDM se beneficia de experiencias acumuladas por las organizaciones participantes en el proyecto, entre las que destacan JISC (Reino Unido) y SURF (Holanda).

La piedra angular del informe *European Landscape Study of Research Data Management*,

publicado en mayo de 2013, gira alrededor de los resultados –desiguales en su cobertura geográfica y representatividad de las respuestas- de los cuestionarios enviados a 4 grandes grupos involucrados en la gestión de datos científicos en Europa (agencias de financiación, organismos nacionales, instituciones de investigación y editores) para definir sus acciones para que la comunidad científica mejore el manejo de sus datos una vez finalizados sus proyectos de investigación. El informe se completa con una introducción bastante exhaustiva que recorre, a partir de buenas prácticas existentes, los elementos que deben formar parte de las políticas y los planes de gestión de datos, y con un par de secciones anejas al análisis de las respuestas a los cuestionarios: por un lado, entrevistas a un grupo de investigadores sobre sus necesidades y un paquete de recomendaciones generales y específicas para cada grupo encuestado.

La sección introductoria ofrece una muy buena panorámica de las claves para diseñar políticas y planes de gestión de datos a partir del estudio de varios documentos de referencia, como el informe *Riding the Wave* (2010), y su continuación *SURFboard for Riding the Wave* (2011), los Principios y Guía de la OCDE, y varios ejemplos concretos de políticas y planes de gestión, principalmente del mundo anglosajón, como son los de la *National Science Foundation* (2011) la *US National Academy of Sciences* (1995) en los Estados Unidos, las recomendaciones del informe de *JISC Dealing with Data* (2007), los principios de políticas de datos de los *Research Councils* en el Reino Unido (2012) y el plan de gestión de la *Australian National University* (2010). Entre las muchas consideraciones, podemos destacar las relativas a los incentivos, a la formación y al apoyo para los investigadores, la inclusión de planes de gestión de datos en la preparación de propuestas, el desarrollo de infraestructuras y de códigos de conducta, la creación de perfiles profesionales específicos, la definición de responsabilidades entre las partes implicadas (autores de datos, agencias financiadoras e instituciones científicas) y las políticas de preservación digital y de utilización de los datos. Una parte especialmente interesante en la introducción se refiere a estudios sobre prácticas en la gestión de datos en disciplinas científicas concretas (p.e, EUDAT) y sobre las posibles reticencias y carencias de la comunidad científica para compartir sus datos (p.e, el proyecto PARSE).

El grueso del informe está compuesto por el análisis de las respuestas de las encuestas sobre políticas de gestión de datos y de las entrevistas a investigadores preparadas por el equipo de SIM4RDM. A pesar de su vocación europea, los resultados de los cuestionarios son relativamente desiguales y en algunos casos –especialmente, la encuesta a editores- han impedido ofrecer conclusiones suficientemente argumentadas. Sin embargo, el ejercicio en su conjunto refleja tendencias generales en materia de planificación de gestión y reutilización de datos a distintos niveles. Las encuestas con el mayor número y variedad de respuestas son las dirigidas a agencias financiadoras y a organismos de investigación. Con respecto a las primeras, cabe destacar el creciente interés y la puesta en marcha de políticas para la financiación y la gestión de datos (casi un 50% de los encuestados), aunque la cifra cae con respecto a la obligatoriedad y la evaluación de planes de gestión de datos en las propuestas de proyectos. En general, se evidencia que la garantía de preservación digital es otra área relativamente poco desarrollada por agencias financiadoras. Por otro lado, el análisis de los resultados de la encuesta sobre políticas institucionales es la más rica en cuestiones y en cobertura geográfica (participación de 15 países) y refleja que más de un tercio de las instituciones encuestadas disponen de una política, si bien suele ser de muy reciente creación. Otro aspecto a resaltar es la precariedad de la mayoría de las muchas iniciativas en marcha para producir herramientas, recursos, guías y apoyo para mejorar la gestión de los datos, y en este sentido quizás pueda establecerse una relación con el bajo número de respuestas recogidas (fundamentalmente de Finlandia, Holanda y Reino Unido) en la encuesta a organismos nacionales encargados de crear códigos de conducta y de coordinar actividades -para el acceso, la formación, la preservación y creación de infraestructuras de datos y el apoyo a agencias financiadoras-.

Entre las principales motivaciones de las instituciones de investigación para desarrollar planes de acción destacan posibles requerimientos legales de agencias financiadoras o institucionales (sobre todo en lo relativo al depósito, gestión y preservación de los datos). La exhaustividad de los planes de gestión varía considerablemente entre las instituciones encuestadas aunque en mayor o menor medida en la mitad de los casos suelen incluir líneas de acción para la captura, mantenimiento y retención de datos, la aplicación de estándares técnicos y bibliográficos, y cuestiones de confidencialidad y de seguridad y de licencias de uso. Por otro lado, el apoyo de las instituciones a la reutilización de los datos suele canalizarse mediante la creación de repositorios, pero los sistemas de incentivos para que los investigadores depositen sus datos siguen siendo minoritarios, con solo un cuarto de las instituciones encuestadas ofreciendo servicios de apoyo como la inclusión de los datos en las listas de publicaciones institucionales, la gestión de datos en la evaluación y auditorías de proyectos, la asignación de DOIs, la traducción de los metadatos al inglés para su mayor visibilidad, la financiación parcial para aplicar formatos estándares etcétera. Las partes menos detalladas del informe corresponden a la encuesta a los editores y las entrevistas a investigadores aunque se dejan entrever cuestiones de relevancia, como la relativa carencia de políticas editoriales para el depósito y el acceso a los datos y la falta de consenso en la citación de los datos y su integración con los artículos correspondientes. Por su parte, las 12 entrevistas a investigadores dejan en evidencia las cuestiones que ellos consideran prioritarias a la hora de gestionar sus datos así como las necesidades más acuciantes. En este sentido, cabe mencionar la existencia de infraestructuras de depósito fáciles de usar, el reconocimiento de la autoría de los datos, el apoyo institucional, los incentivos para su depósito y formación.

Por último, el informe cierra con una serie de conclusiones generales y recomendaciones para que los organismos nacionales, agencias financiadoras, instituciones de investigación y editores promuevan una mejor gestión de datos por parte de la comunidad científica. El trabajo futuro de SIM4RDM (en vigor hasta la primavera de 2014) prevé la inclusión de más stakeholders en el ámbito de los datos científicos, como por ejemplo más órganos nacionales y organizaciones internacionales, centros de datos y proveedores de infraestructuras (bibliotecas), comités editoriales de revistas científicas y sociedades académicas.

Informes como éste son una prueba del interés que despierta este tema, y están alentados en gran medida por el auge del acceso abierto, la multiplicación de mandatos sobre la difusión de los resultados de investigación y por los desarrollos tecnológicos de última hora. El impulso de las agencias financiadoras es clave para cambiar de modo fundamental las prácticas en la gestión y la comunicación de estos resultados de investigación y las Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020, recién publicadas a mitad de diciembre, tendrán sin duda un efecto positivo en el panorama europeo. Sin embargo, existen varios interrogantes en torno a la gestión de este trozo muy considerable de ciencia, empezando por la necesidad de abordar desde un consenso internacional una definición estándar de lo que entendemos por “datos de investigación” y no repetir los errores derivados de la relativa falta de armonización entre los múltiples modelos de mandatos de acceso abierto a las publicaciones científicas —en este sentido, la aplicación de estándares internacionales y la identificación de excepciones son importantes. Por otra parte, la creciente facilidad tecnológica para recuperar información sobre su impacto y el tráfico web que generan, para analizarlos mediante técnicas de minería y las perspectivas de integrarlos en ejercicios de evaluación ha abierto la puerta a nuevas iniciativas con una vocación fuertemente comercial (bases de datos bibliográficas, repositorios de pago, servicios de alojamiento masivo etc) que podrían tener un efecto no deseado sobre el objetivo final de mayor descubrimiento, apertura y reutilización en aras del progreso científico y socioeconómico. Otras cuestiones que quedan para trabajo futuro se refieren a la siempre delicada cuestión de la gestión y preservación de la integridad de los datos, la protección de privacidad y las dificultades de atribuir adecuadamente estas obras,

de fuerte impronta colectiva en muchas disciplinas, a sus correspondientes autores.

**URL de origen:** <http://www.ub.edu/blokdebid/es/content/alud-de-datos>

**Enlaces:**

[1] <http://www.ub.edu/blokdebid/es/users/isabel-bernal>

[2] <http://www.ub.edu/blokdebid/es/category/bullet%C3%AD/blok-de-bid-esp>

[3] <http://www.ub.edu/blokdebid/es/category/etiquetes/datos-de-investigaci%C3%B3n>

[4] <http://www.ub.edu/blokdebid/es/category/ressenyadors/isabel-bernal>

[5] <http://www.ub.edu/blokdebid/es/category/etiquetes/politicas>

[6] [http://www.sim4rdm.eu/sites/default/files/uploads/documents/SIM4RDM%20landscape%20report%20vs1%204\\_14.08.13.pdf](http://www.sim4rdm.eu/sites/default/files/uploads/documents/SIM4RDM%20landscape%20report%20vs1%204_14.08.13.pdf)

[7] <http://www.sim4rdm.eu/>

[8] [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)