

# Introducción

La era de los datos ya está aquí. Ante la avalancha de información, la sociedad al completo necesita datos con los que corroborar de forma concreta lo que se toma como cierto. Diríamos que esta forma de interpretar la realidad está bastante cerca de lo que hacían los científicos: analizar las evidencias que se han desarrollado con un método, que se explicita para permitir la comprobación de las hipótesis. Esta forma de trabajar ha llegado incluso a los niveles más altos de los gobiernos. La Cumbre de Líderes del G8 (junio del 2013) incluye como valiosos para la economía los datos derivados de la investigación y los resultados de experimentos científicos. Pero, ¿por dónde empezar a abordar un panorama tan complejo? Es el momento de esbozar una hoja de ruta que permita asesorar a los investigadores y a los propios organismos de investigación. Las buenas prácticas de otros países pueden guiar nuestros pasos: Estados Unidos, Australia, Gran Bretaña.

Este libro responde a preguntas como: ¿Qué son los datos de investigación? ¿Por qué debo compartirlos? ¿Qué herramientas de cita existen? ¿Qué es *open access* y ciencia abierta? ¿Cómo almaceno mis datos? ¿Qué metadatos debo conocer? ¿Qué datos han de preservarse a largo plazo y cuál es la función de una biblioteca en ello? ¿Qué son los requerimientos de los financiadores sobre los datos? ¿Cómo descubro los datos que los investigadores están creando? ¿Qué recursos existen para trabajar con la gestión de datos?

A lo largo del texto se desarrollan los contenidos para conocer el tema de gestión de los datos en profundidad y se proponen mediante recuadros posibles ampliaciones con el objetivo de guiar nuevos descubrimientos. El mapa de las infraestructuras que presentamos incluye no solo al personal involucrado o la tecnología, sino también las políticas o los estándares habituales. Esta multitud de recursos puede ser utilizada en diferentes contextos, tanto por los grupos de investigación, los gestores de la ciencia y los infomediarios como por bibliotecarios o documentalistas.

El trabajo se estructura en cuatro capítulos. El primero de ellos encuadra la gestión de los datos en el contexto de la ciencia abierta. Su primer apartado, «Agentes sociales involucrados y sus roles», pretende reflejar la complejidad del ecosistema de la ciencia actual. El segundo, «Actividades científicas», da a conocer cuál es el trabajo principal de los investigadores actualmente. El tercer apartado da cuenta de las «Nuevas formas de divulgar la ciencia» y algunas de las herramientas digitales

que pueden ayudar en la investigación como Many Eyes o Galaxy Zoo. Su cuarto punto, «*Research data*», centra los conceptos que se han barajado hasta entonces en los datos derivados de la investigación, y también ofrece información sobre cómo está la situación actualmente según algunas encuestas nacionales e internacionales.

El segundo capítulo muestra ya el corazón de los contenidos. En él se atienden los aspectos necesarios para desarrollar el trabajo con datos y se revisan los métodos de la investigación, de la planificación del ciclo de vida de la información y el papel que un gestor de información asume con respecto a ellos. Su primer apartado, «Producción de datos de un proyecto de investigación», desarrolla aspectos que quizá no son muy conocidos por la comunidad de gestión de la información y que son absolutamente imprescindibles para poder generar servicios de apoyo. Entender el proceso investigador facilita la adaptación de los perfiles profesionales al nuevo entorno de trabajo con datos. Su segundo punto, la «Perspectiva de las agencias de financiación», comenta los casos de estudio británico, estadounidense y europeo en cuanto a los requerimientos de las agencias financiadoras. El tercero, sobre la gestión del ciclo de vida de los datos en una organización, aborda la planificación de un servicio de gestión de datos integral, incidiendo en los aspectos de gestión que son inherentes al trabajo con datos. Su cuarto apartado, «Papel del *data manager*», contiene información útil para crear un programa de formación adecuado a los retos a los que actualmente se enfrenta un gestor de información.

El tercer capítulo cubre los contenidos que es necesario conocer para trabajar de forma global en el sector de los datos. Su primer apartado, sobre fuentes de datos, resume cómo ahora mismo se están depositando los datos y relata los proyectos de tipo metodológico. El segundo, sobre infraestructuras, clasifica y recoge las plataformas técnicas que existen. El tercer punto, los «Casos de éxito», detalla dos experiencias muy distintas: una de alcance nacional, ANDS (Australian National Data System), y otra de alcance mundial, RDA (Research Data Alliance).

El último capítulo concluye con las lecciones aprendidas hasta el momento a escala internacional. Sus apartados, sobre pautas y guías y tendencias recientes, pretenden ser un resumen de la cuestión y ofrecer recursos útiles al lector.

Este libro ha podido desarrollarse gracias al apoyo en la investigación por parte del Plan Nacional I+D y la Fundación MAPFRE «OpenDataScience. Centro de recursos para la preservación y gestión de datos abiertos de investigación», CS02012-39632-C02-02, <<http://www.datasea.es>>. Tiene sus orígenes en un curso impartido en SEDIC en marzo del 2014.