

# Bibliotecas y datos: gestionando los bienes científicos para el futuro

Mirelys Puerta-Díaz<sup>1</sup>, Paloma Marín-Arraiza<sup>2</sup>, Silvana Aparecida Borsetti Gregorio Vidotti<sup>3</sup>

<sup>1</sup> <https://orcid.org/0000-0002-2312-2540>. Universidade Estadual Paulista -UNESP (Marília, São Paulo, Brasil). [mirelys.puerta@unesp.br](mailto:mirelys.puerta@unesp.br)

<sup>2</sup> <https://orcid.org/0000-0001-7460-7794>. Universidade Estadual Paulista -UNESP (Marília, São Paulo, Brasil) y TU Wien Bibliothek (Viena, Austria). [paloma.arraiza@unesp.br](mailto:paloma.arraiza@unesp.br), [paloma.arraiza@tuwien.ac.at](mailto:paloma.arraiza@tuwien.ac.at)

<sup>3</sup> <https://orcid.org/0000-0002-4216-0374>. Universidade Estadual Paulista -UNESP (Marília, São Paulo, Brasil). [silvana.vidotti@unesp.br](mailto:silvana.vidotti@unesp.br)

**Tipo de contribución:** comunicación oral.

**Palabras clave:** interoperabilidad de datos; servicios bibliotecarios; bibliotecario de datos; identificadores persistentes; gestión de datos de investigación.

## 1 Introducción

Alrededor de la década de 1990 surgen los primeros repositorios de documentos digitales. Desde ese momento la interoperabilidad de los contenidos en internet empieza a adquirir relevancia. Así en 2001 se publica el *Open Archive Initiative-Protocol for Metadata Harvesting* (OAI-PMH) como iniciativa para garantizar esta interoperabilidad. El protocolo OAI-PMH es actualmente uno de los más utilizados en repositorios y otros sistemas de información y para permitir la compartición de datos entre sistemas (Bardi & Mangui, 2015).

En la actualidad, la interoperabilidad y preservación de datos está todavía más presente, dado el considerable cambio que viene experimentando la práctica científica en los últimos años. Podemos hablar de un paradigma científico basado en datos (Hey, Tansley & Tolle, 2009). En él, los datos de investigación desempeñan un papel fundamental a lo largo del proceso científico, desde la fase de planificación hasta la publicación de los resultados de investigación. Puesto que los datos son cruciales en procesos como los controles de reproducibilidad, innovación y toma de decisiones, representan un importante e inestimable capital intelectual dentro del ecosistema de investigación (Kowalczyk & Shankar, 2011; Mons, 2018).

Es importante destacar que con el término «datos de investigación» se hace referencia a todo tipo de información (sin importar su formato) necesaria para sostener y validar el proceso científico. En los datos de investigación se incluyen todos los materiales creados en el transcurso del trabajo académico, por ejemplo, mediante digitalizaciones, grabaciones, capturas fotográficas, experimentos, mediciones, encuestas y entrevistas. Para garantizar su reproducibilidad, los datos deben estar en conformidad con los principios FAIR (encontrable, accesible, interoperable, reutilizable, del inglés *findable, accessible, interoperable, reusable*). Estos principios se establecieron para proporcionar unas directrices y medidas para la correcta gestión de datos. Además, no solo se aplican a los datos, sino también a los sistemas de información (Wilkinson *et al.*, 2016).

Si bien la responsabilidad por la gestión de datos corresponde en primera instancia a los investigadores, resulta necesario que las bibliotecas universitarias y de investigación amplíen sus servicios para adaptarse a las nuevas demandas de los investigadores con respecto a la gestión de datos. Al implementar nuevos servicios se persigue la consecución de una mejor y más cuidadosa

gestión y custodia de datos (*data stewardship*) y de un apoyo directo a las tareas investigativas a lo largo de todo su ciclo.

Las buenas prácticas en la gestión de datos garantizan, que más allá de los propósitos para los que fueron creados los datos originalmente, puedan ser reutilizados para el descubrimiento de nuevo conocimiento, hecho de trascendental importancia para la ciencia abierta (Plotkin, 2014; Mons, 2018).

En este contexto surge también la figura del bibliotecario de datos (*data librarian*) como agente propulsor de nuevos servicios bibliotecarios orientados a la preservación de los datos de investigación, garantizar la sostenibilidad de los sistemas de almacenamiento y la reusabilidad de los datos. La reusabilidad queda garantizada cuando se produce una transferencia de los datos entre experimentos (Barba, 2018).

## 2 Procedimientos metodológicos

La metodología de este trabajo es cualitativa y se encaja dentro de una investigación exploratoria que pretende dar una visión general sobre el tratamiento de los datos dentro de las bibliotecas.

Junto con el análisis bibliográfico realizado en diferentes bases de datos, tales como *Scopus*, *Web of Science* y *Dimensions*, se realiza un mapeo de los servicios de diferentes bibliotecas universitarias y de investigación europeas y latinoamericanas. Para este mapeo, se parte del ciclo de vida de la investigación propuesto por JISC (anteriormente *Joint Information Systems Committee*) y se identifican servicios que las bibliotecas prestan en cada etapa a partir de los servicios que aparecen en sus páginas web. Al unir estas dos etapas se establece el marco teórico del trabajo del bibliotecario de datos (*data librarian*) y el administrador o gestor especialista de datos (*data steward*), así como la actuación de las bibliotecas en la gestión de los datos como patrimonio científico para el futuro.

De esta forma se observan las tendencias de trabajo y los nuevos desarrollos en este área y se identifican nuevas líneas de trabajo y pautas que podrían ser implementadas en el ámbito bibliotecario, especialmente en bibliotecas universitarias y de investigación.

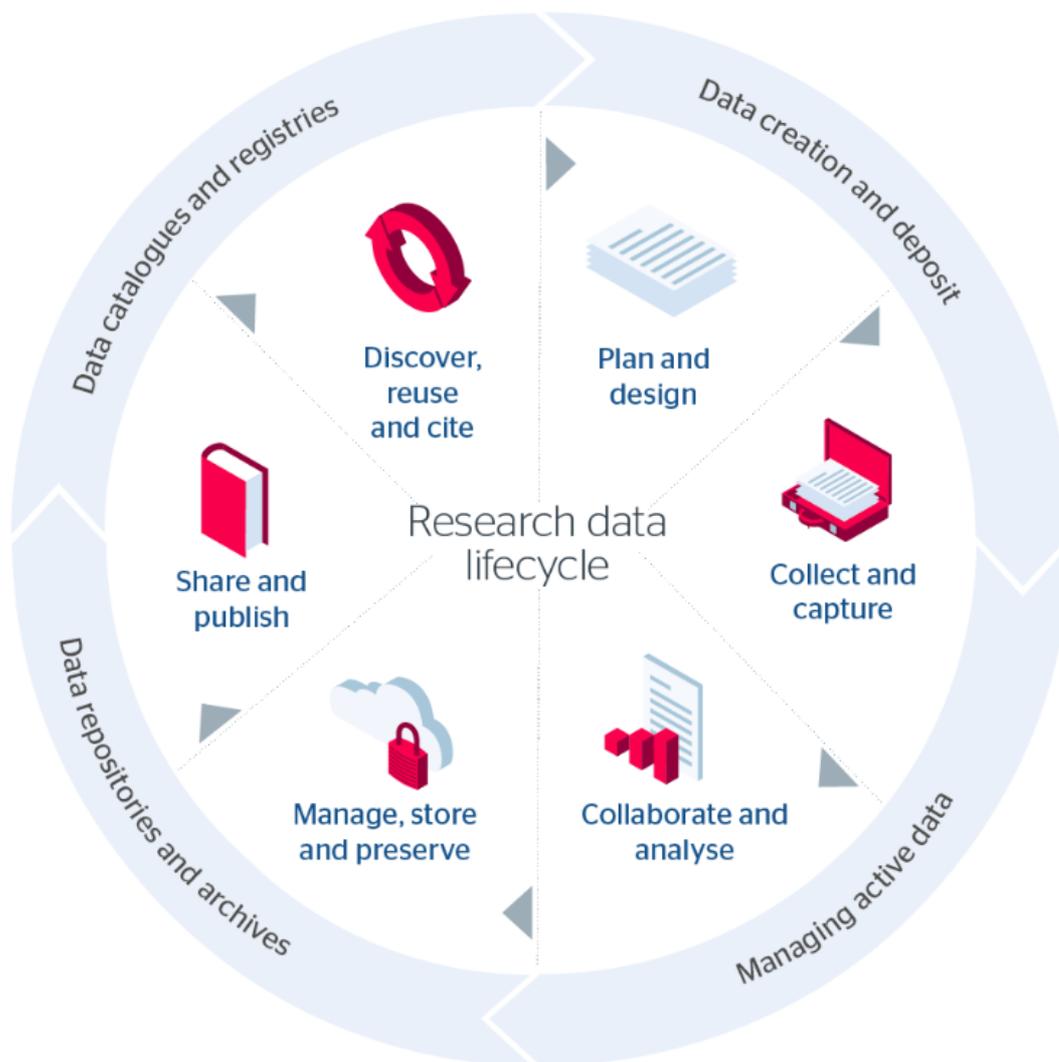
## 3 Resultados principales: Servicios bibliotecarios para el tratamiento de datos de investigación

El tratamiento de datos y metadatos, la información descriptiva y técnica sobre los datos, es central en el ciclo de investigación, tanto en las fases de planificación y creación como en la fase de publicación y preservación, ya que permite identificar el tipo de proyecto y su contexto. Los datos de investigación se pueden considerar bienes que deben preservarse para el futuro.

Las bibliotecas desempeñan un papel muy importante en este contexto de preservación. Podemos ver su participación directa en el desarrollo de materiales de formación sobre buenas prácticas para la apertura de los resultados de investigación y para garantizar que la Ciencia Abierta se convierta en el *modus operandi* (por ejemplo, el *toolkit* del proyecto FOSTER Open Science, <https://www.fosteropenscience.eu/toolkit>) o en el creación y curación de modelos de datos específicos (por ejemplo, el modelo de datos de Europea).

Otra forma de observar cómo se realiza ese tratamiento de datos es mediante las representaciones gráficas de los ciclos de vida datos. Un ejemplo de estas representaciones se ve en la figura 1.

Figura 1. Ciclo de vida de los datos de investigación. Fuente: JISC, 2016. Licencia: CC-BY-ND.



En la primera fase del ciclo (creación y depósito de datos) representado en la figura, los planes de gestión de datos desempeñan un papel muy importante y está siendo solicitados por muchas agencias de financiación a nivel mundial. Por ejemplo, para los proyectos de Horizon 2020 o aquellos financiados por las diferentes Fundaciones Estatales de Amparo a la Investigación de Brasil.

En la tercera fase (repositorios y archivos de datos), los servicios de repositorios, CRIS y servicios institucionales de almacenamiento en la nube son especialmente relevantes para la gestión de los datos durante la realización del proyecto y tras haber realizado el proyecto. Es importante distinguir aquí entre los datos compartidos durante el proyecto y los que finalmente se publican.

La cuarta fase (catálogos y registros de datos) se caracteriza por la publicación, descubrimiento y reuso y citación de datos. En esta fase, los identificadores persistentes desempeñan una función principal.

Además de los servicios específicos para cada fase del ciclo, se ha incorporado al contexto de la gestión de datos la figura del bibliotecario de datos (*data librarian*) y de forma incluso más específica, la figura del gestor especialista de datos (*data steward*).

A continuación, se presentan los resultados de un primer mapeo sobre servicios para la preservación e interoperabilidad de datos de investigación ofrecidos por parte de bibliotecas europeas y latinoamericanas.

### 3.1 Herramientas para la generación de planes de gestión de datos

Dada la diversidad de datos científicos, su naturaleza heterogénea y las características de la cultura de cada comunidad científica, se considera una buena práctica institucional proporcionar un mecanismo a los investigadores que les indique cómo administrar los datos asociados al ciclo de vida de la investigación (FECYT, 2012). Los planes de gestión de datos constituyen un mecanismo adecuado para ahorrar tiempo y esfuerzo en el proceso de la investigación, así como potenciar la productividad y visibilidad del investigador (Smale *et al.*, 2018). Por lo tanto, se han convertido en un elemento clave dentro de la gestión de datos.

Los planes de gestión de datos son documentos vivos que describen la gestión de los datos a lo largo de su ciclo de vida. Según la Comisión Europea, un plan de gestión de datos debe incluir la siguiente información: (1) el tratamiento de los datos de investigación durante y tras el proyecto; (2) el tipo de datos que será colectado, procesado y/o generado; (3) la metodología y estándares que se utilizarán; (4) si los datos se compartirán o disponibilizarán en acceso abierto; (5) cómo se realizará la curaduría y preservación de los datos.

Sin embargo, la creación de planes de datos se observa entre la comunidad científica como una actividad que consume mucho tiempo (Williams, Bagwell & Nahm Zozus, 2017). Por eso, han surgido algunas herramientas online para la generación de planes de gestión de datos como DMPonline del Digital Curation Centre (<https://dmponline.dcc.ac.uk/>) o DMPTool de la UC3 California Digital Library (<https://dmp.cdlib.org/>).

En el contexto europeo, destacamos la iniciativa de la biblioteca universitaria de la Universidad de Helsinki DMPTuuli (<https://www.dmptuuli.fi/>). Esta herramienta de automatización de la creación de planes de gestión de datos cuenta con más de 7000 usuarios procedentes de más de 40 instituciones finlandesas. Las actividades de conceptualización y coordinación de DMPTuuli forman parte del trabajo de la biblioteca.

Asimismo, desde la Research Data Alliance (RDA), en la que bibliotecas universitarias y de investigación también participan activamente, abogan por el uso de planes de gestión de datos legibles por máquinas, que permitirían integrar estos planes en los flujos de trabajo de las diferentes partes interesadas del ecosistema de investigación (Miska, *et al.* 2019). Entre otros, en esas partes interesadas están los gestores de repositorios que, en muchos casos, son las bibliotecas universitarias.

### 3.2 Repositorios, repositorios de datos y sistemas CRIS institucionales

Diversas universidades disponibilizan algún tipo de sistema de información para preservar y difundir los resultados de investigación de la institución. Además, estamos presenciando un fuerte desarrollo de políticas públicas, cuyo objetivo es la apertura y preservación de los resultados de aquellos proyectos financiados con fondos públicos. Un ejemplo se ve en la Resolución 753 - E/2016 del Ministerio de Ciencia Argentino que reglamenta la ley de creación de repositorios digitales institucionales de acceso abierto.

En el continente americano, Brasil se caracteriza por su desarrollo en acceso abierto al conocimiento científico (Rodrigues & Abadal, 2014), lo que se contempla también en el amplio número de repositorios institucionales y de datos que disponibiliza (cien según el dato de OpenDOAR en mayo de 2019). Dentro del Estado de San Pablo se incide mucho en la interoperabilidad y se disponibiliza la búsqueda integrada entre los tres repositorios de las universidades públicas del estado. Asimismo, dentro del departamento de Ciência da Informação da Faculdade de Filosofia e Ciências da Universidade Estadual Paulista (UNESP) se inició un proyecto piloto de sistema CRIS para los grupos de investigación del área (Vidotti *et al.*, 2017).

Por su parte, en España, el equipo a cargo de Digital.CSIC del Consejo Superior de Investigaciones Científicas (<https://digital.csic.es/>) colabora en el desarrollo del proyecto «Infraestructuras y Estándares para la Ciencia en Abierto» (INEOS). Otro ejemplo se encuentra en Austria, donde las bibliotecas universitarias de diferentes instituciones de enseñanza superior colaboran en la implementación de infraestructuras para *e-Science* dentro del proyecto nacional «e-Infrastructures Austria Plus» (<https://www.e-infrastructures.at/en/>). Otro objetivo de este proyecto es institucionalizar la gestión de datos y establecer especialistas en los conceptos FAIR.

### 3.3 Servicios de asignación de identificadores persistentes

Los identificadores persistentes son uno de los puntos principales para garantizar la encontrabilidad de los datos, como se refleja en el principio F1 de los principios FAIR (<https://www.go-fair.org/fair-principles/>): «A los (meta)datos se les asigna un identificador persistente globalmente único». También el proyecto europeo FREYA (<https://www.project-freya.eu/en/about/mission>) tiene entre sus objetivos que los identificadores persistentes formen parte de la investigación abierta. Estos identificadores pueden asignarse a agentes, recursos, declaraciones de derechos, eventos o entidades derivadas (Dappert *et al.*, 2017). Pueden permitir la conexión e interoperabilidad de todas las entidades, aumentando la encontrabilidad de objetos nativos del contexto digital que no han sido interconectados con sus creadores (Klein & van der Sompel, 2017).

En este contexto, varias bibliotecas de investigación europeas se han convertido en DOI-*National Desk*, es decir, en centro de asignación de identificadores de objeto digital (DOI, su sigla en inglés)—los identificadores persistentes de recursos electrónicos—. Como ejemplo tenemos la Biblioteca Nacional Alemana de Ciencia y Tecnología (<https://www.tib.eu/en/publishing-archiving/doi-service/>) y a la Biblioteca Universitaria de la universidad suiza ETH Zürich (<http://www.library.ethz.ch/en/Services/Publishing-registering-managing/ETH-Zurich-DOI-Desk>). Asimismo, en Alemania el consorcio de ORCID (Identificador Abierto de Investigador y Colaborador) trabaja con la Biblioteca Nacional Alemana para enlazar los identificadores persistentes personales ORCID iD con los registros del control de autoridades *Gemeinsame Normdatei* (GND). Por su parte, el consorcio brasileño de ORCID está compuesto, entre otros, por la biblioteca científica SciELO, cuyo objetivo será la integración del sistema ORCID en la biblioteca en línea para contribuir a una mayor visibilidad de las investigaciones.

### 3.4 Integración de la figura *Data steward*

La integración de la figura del administrador o gestor especialista de datos (*data steward*) es un avance en las bibliotecas universitarias para la gestión de datos. Los administradores de datos se establecen de forma similar a los bibliotecarios especializados (*liaison librarians*) pero concentrándose en la gestión de datos de investigación.

Kennan (2016) enuncia las competencias deseadas para los bibliotecarios de datos, a saber: (1) habilidades interpersonales, características de comportamiento y habilidades comunicativas; (2) dominio del ecosistema de investigación; (3) conocimiento específico sobre el uso de datos; y (4) conocimiento de las tecnologías de la información. Un administrador de datos debe unir el conocimiento sobre gestión de datos propios de un bibliotecario de datos y el conocimiento específico de cada área científica.

La biblioteca de la Universidad Técnica de Delft en los Países Bajos ha sido pionera en la integración de esta figura (<https://www.tudelft.nl/en/library/current-topics/research-data-management/research-data-management/data-stewardship/>). Entre sus tareas se encuentran poner en

marcha almacenamientos seguros para datos, asesorar en buenas prácticas para la gestión de datos, ofrecer información sobre el archivo de datos, dar consejos sobre cómo aumentar el impacto a través de la compartición de datos, ayudar en la preparación de planes de gestión de datos.

#### 4 Consideraciones parciales

Como apuntan Mackenzie y Martin (2016) las bibliotecas universitarias y de investigación se enfrentan a un nuevo paradigma de datos que ha cambiado la práctica científica y su sistematización. Esta transformación implica un cambio de roles, nuevos roles y la necesidad de desarrollar nuevas capacidades por el bibliotecario, hacia la integración de las figuras del bibliotecario de datos y del *data steward*. Estos profesionales, con un gran conocimiento sobre el ecosistema de investigación, las tecnologías de información y las fuentes de datos de cada área, asistirían a los investigadores en la consulta y publicación de trabajos en abierto y la necesaria gestión de datos de investigación.

De esta forma, las bibliotecas pueden pasar de un modelo basado en servicios a uno basado en coparticipación y colaboración con los investigadores. Concretamente las bibliotecas desempeñarían un papel más proactivo en el proceso de creación de conocimiento y se ubicarían en cada fase del proceso de planificación investigativa, creando estrategias para su preservación a largo plazo.

Estos cambios se han podido observar en este mapeo inicial de servicios bibliotecarios, cuyo objetivo debe ser centrarse en la demandas emergentes para la preservación de los resultados de investigación. De esta forma algunas bibliotecas, del área europea y latinoamericana, ofrecen servicios orientados a la asignación de identificadores persistentes, el desarrollo de sistemas CRIS institucionales y herramientas para la generación de planes de gestión de datos de investigación.

Como siguientes pasos, consideramos analizar en profundidad esos servicios y extender el mapeo. Además, consideramos importante ver los pasos que son necesarios para el establecimiento de estos servicios. Una aproximación inicial sería la creación de una política institucional de gestión de datos de investigación, seguida de la integración de un centro de gestión de datos de investigación y, por último, la integración específica de bibliotecarios de datos y/o *data stewards* que puedan ofrecer servicios de asesoría ajustados a los perfiles específicos de los investigadores de cada área.

#### 5 Referencias

Barba, L. A. (2018). Terminologies for reproducible research. *Preprint* en: [arXiv:1802.03311](https://arxiv.org/abs/1802.03311)

Bardi, A., & Manghi, P. (2015). A Framework Supporting the Shift from Traditional Digital Publications to Enhanced Publications. *D-Lib Magazine*, 21(1/2). <https://doi.org/10.1045/january2015-bardi>

Comisión Europea. Data Management. Disponible en: [http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management\\_en.htm](http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm)

Dappert, A, *et al.* (2017). Connecting the Persistent Identifier Ecosystem: Building the Technical

and Human Infrastructure for Open Research. *Data Science Journal*, 16(28), 1–16. DOI: <https://doi.org/10.5334/dsj-2017-028>

Europeana. Europeana Data Model. Disponible en: <https://pro.europeana.eu/resources/standardization-tools/edm-documentation>

FOSTER Open Science. Open Science Training Courses <https://www.fosteropenscience.eu/toolkit>

Hey, A. J., Tansley, S., & Tolle, K. (2009). *The fourth paradigm: Data-intensive scientific discovery*. Redmond, WA: Microsoft Research. Disponible en: [https://www.microsoft.com/en-us/research/wp-content/uploads/2009/10/Fourth\\_Paradigm.pdf](https://www.microsoft.com/en-us/research/wp-content/uploads/2009/10/Fourth_Paradigm.pdf)

Kennan, Mary Anne. (2016). *Data Management: Knowledge and skills required in research, scientific and technical organisations*. Paper presented at the IFLA WLIC 2016 – Columbus, OH – Connections. Collaboration. Community in Session 221 - Science and Technology. Disponible en: <http://library.ifla.org/id/eprint/1466>.

Klein, M., and Van de Sompel, H. (2017) Discovering Scholarly Orphans Using ORCID. Proceedings of the 2017 ACM/IEEE Joint Conference on Digital Libraries. Arxiv preprint arXiv:1703.09343 en <https://arxiv.org/abs/1703.09343>

Kowalczyk, S., & Shankar, K. (2011). Data sharing in the sciences. *Annual review of information science and technology*, 45(1), 247-294. <https://doi.org/10.1002/aris.2011.1440450113>

Mackenzie, A. & Martin, L. (2016). *Developing Digital Scholarship. Emerging practices in academic libraries*. London: Facet Publishing.

Miksa, T., Simms, S., Mietchen, D., & Jones, S. (2019). Ten principles for machine-actionable data management plans. *PLOS Computational Biology*, 15(3), e1006750. <https://doi.org/10.1371/journal.pcbi.1006750>

Mons, B. (2018). *Data Stewardship for Open Science*. Chapman and Hall/CRC. <https://doi.org/10.1201/9781315380711>

Plotkin, D. (2014) *Data stewardship: An actionable guide to effective data management and data governance*. Boston: Morgan Kaufmann. ISBN:978-0-12-410389-4. pp 231.

Rice, R. & Southall, J. (2016). *The Data Librarian's Handbook*. London: Facet Publishing.

Rodrigues, R. S., & Abadal, E. (2014). Scientific journals in Brazil and Spain: Alternative publishing models. *Journal of the Association for Information Science and Technology*, 65(10), 2145-2151. <https://doi.org/10.1002/asi.23115>

Smale, N., Unsworth, K. J., Denyer, G., & Barr, D. P. (2018). *The History, Advocacy and Efficacy*

of Data Management Plans. *bioRxiv*, 443499. Disponible en:  
<https://www.biorxiv.org/content/early/2018/10/17/443499>

Vidotti, S. A., Coneglian, C. S., Roa-Martínez, S., Arakaki, F., Brandt, M. B., & Da Costa, A. J. (2017). Repositório de dados de pesquisa para grupo de pesquisa: Um projeto piloto. *Informação & Tecnologia (ITEC)*, 4(2), 221-242. Disponible en:  
<http://www.periodicos.ufpb.br/ojs/index.php/itec/article/view/40198>

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>

Williams, M., Bagwell, J., & Nahm Zozus, M. (2017). Data management plans: the missing perspective. *Journal of Biomedical Informatics*, 71, 130-142.  
<https://doi.org/10.1016/j.jbi.2017.05.004>