

ANÁLISIS DE LOS PRINCIPIOS PARA LA DESCRIPCIÓN DE DATOS DE INVESTIGACIÓN MEDIANTE DATACITE METADATA SCHEMA

*Eder Ávila Barrientos**

Universidad Nacional Autónoma de México. Instituto de Investigaciones Bibliotecológicas y de la Información.

Resumen: El objetivo de este trabajo consiste en analizar los principios teórico-metodológicos relacionados con la descripción de los datos de investigación. Se realizó un análisis sobre el estado de la cuestión de los datos de investigación, en el cual se abordan aspectos de su citación, descripción y sistematización. Se identificaron y analizaron los elementos de metadatos para la descripción de conjuntos de datos de investigación que se incluyen en el DataCite Metadata Schema, con el propósito de crear una propuesta de perfil descriptivo aplicable a estos conjuntos. Se estima que, si los datos de investigación se encuentran debidamente descritos, entonces se fomentará en mayor grado su accesibilidad y reutilización. Para ello, es necesario que las instituciones académicas y de investigación participen en la generación de políticas de acceso abierto a sus datos de investigación.

Palabras clave: datos de investigación; descripción; recuperación; repositorio de datos.

Title: ANALYSIS OF THE PRINCIPLES FOR THE DESCRIPTION OF RESEARCH DATA BY DATACITE METADATA SCHEMA.

Abstract: The objective of this work is to analyze the theoretical-methodological principles related to the description of the research data. An analysis was carried out on the status of the research data issue, which addresses aspects of its citation, description and systematization. Metadata elements for the description of research data sets that are included in the Metadata Schema DataCite were identified and analyzed, with the purpose of creating a proposed descriptive profile applicable to these sets. It is estimated that, if the research data is properly described, then its accessibility and reuse will be further promoted. To do this, it is necessary for academic and research institutions to participate in the generation of open access policies for their research data.

Keywords: research data; description; retrieval; data repository.

Copyright: © 2020 Servicio de Publicaciones de la Universidad de Murcia (Spain). Este es un artículo de acceso abierto distribuido bajo los términos de la licencia Creative Commons Reconocimiento 4.0 Internacional (CC BY 4.0).

1 INTRODUCCIÓN

El siglo XXI se caracteriza por la revolución y explosión de los datos en los diversos contextos de la actividad humana. Diariamente se convive con datos que registran el comportamiento de la realidad y que permiten comprender los diversos fenómenos que acontecen en ella. Al respecto Harari (2015, p. 428) plantea la siguiente pregunta: ¿acaso hay algo en el universo que no pueda reducirse a datos? Las tecnologías de la información y comunicación han contribuido a la conformación de un contexto *datificado* y al surgimiento de nuevos fenómenos y problemáticas relacionadas con su uso. No obstante, el ser humano siempre se encuentra generando, compartiendo y consumiendo datos. En el ámbito científico y académico, los datos forman parte del proceso natural de la búsqueda de nuevos conocimientos. Pues, “la ciencia siempre se ha basado en la utilización de datos, pero hasta hace poco, estos datos de investigación eran difíciles de recoger, conservar, compartir y reutilizar” (CSUC, 2019, p. 7). Los datos de investigación (DI) son el resultado de experimentos, procesos y metodologías que son aplicados en el contexto de la investigación científica, académica, pura y aplicada. Son la base de los resultados que se obtienen como parte de un procedimiento indagatorio. Muchos de estos datos se encuentran estructurados y no estructurados y su accesibilidad puede ser abierta o cerrada en concordancia con las características de la comunidad en donde se generan.

En el marco de la investigación, los datos sirven para representar fenómenos. Pues, “siempre han sido la base del progreso científico, aunque hasta hace poco ha sido difícil compartirllos y analizarlos. Internet ha brindado la oportunidad de transmitir conjuntos de datos grandes (y pequeños) de manera rápida y fácil, y ponerlos a disposición de cualquiera” (Callaghan *et al.*, 2013, p. 194).

Simultáneamente, la web hace posible consultar plataformas que registran y almacenan conjuntos de datos de investigación. Por ejemplo, el Registry of Research Data Repositories (<https://www.re3data.org/>) “un registro global de repositorios de datos de investigación” (R3data, párr. 1). En esta fuente se reúnen repositorios que remiten a

* eder@iibi.unam.mx

conjuntos de datos de investigación de diferentes disciplinas académicas. Por consiguiente, la generación de datos en los diferentes contextos disciplinarios permite visualizar un escenario en donde existen conjuntos de datos de investigación de diversa temática y naturaleza. Estos factores dejan entrever la necesidad de contar con principios que fomenten la gestión de dichos datos, en los cuales se contemplen pautas para su descripción, citación y visibilidad.

Se estima que, si los datos de investigación se encuentran debidamente descritos y representados, entonces se fomentará en mayor grado la confiabilidad de su reutilización. De ahí la necesidad de seguir principios normalizados e internacionales para alcanzar este propósito.

Bajo esta premisa, se encuentran The Data FAIR Principles, los cuales “hacen énfasis en mejorar la capacidad de las computadoras para encontrar y utilizar automáticamente los datos, además de fomentar su reutilización por parte de las personas” (Wilkinson, 2016, p. 1). Estos principios acumulables se explican de la siguiente manera:

<p>Findability=Encontrabilidad <i>Para encontrar a los datos:</i> F1. A los (meta) datos se les asigna un identificador globalmente único y persistente; F2. Los datos se describen con metadatos enriquecidos; F3. Los metadatos incluyen clara y explícitamente al identificador de los datos que describe; F4. Los (meta) datos se registran o indexan en un recurso de búsqueda.</p>
<p>Accessibility=Accesibilidad <i>Para acceder a los datos:</i> A1. Los (meta) datos son recuperables por su identificador utilizando un protocolo estandarizado de comunicación; A1.1. El protocolo debe ser abierto, gratuito y universalmente implementable; A1.2. El protocolo permite efectuar un procedimiento de autenticación y autorización, cuando sea necesario; A2. Los metadatos son accesibles, incluso cuando los datos ya no están disponibles.</p>
<p>Interoperability=Interoperabilidad <i>Para hacer interoperables a los datos:</i> I1 Los (meta) datos usan un lenguaje formal, accesible, compartido y ampliamente aplicable para la representación del conocimiento; I2 Los (meta) datos usan vocabularios que siguen los principios FAIR; I3 Los (meta) datos incluyen referencias calificadas a otros (meta) datos.</p>
<p>Reusability=Reusabilidad <i>Para reutilizar los datos:</i> R1. Los meta (datos) se describen detalladamente con una pluralidad de atributos precisos y relevantes; R1.1. Los (meta) datos se liberan con una licencia de uso de datos clara y accesible; R1.2. Los (meta) datos están asociados a una procedencia detallada; R1.3. Los (meta) datos cumplen con estándares relevantes para la comunidad y el dominio en donde se generan.</p>

Tabla I. The Data FAIR Principles. Fuente: elaboración propia. Con datos de (Wilkinson, 2016, p. 4).

En la Tabla I se observa que el principio de encontrabilidad hace referencia a la utilización de identificadores únicos y persistentes que permitan buscar y encontrar a los conjuntos de datos de investigación. A su vez, la accesibilidad a los datos contempla su recuperación mediante la asignación de un identificador universal, estandarizado y abierto. Además, su interoperabilidad se establece mediante la utilización de lenguajes y vocabularios que hagan posible representar formal y ampliamente sus atributos, es decir, referenciar a los datos mediante un esquema de representación común y compatible entre sistemas. Finalmente, la reusabilidad es un principio que se establece para describir de manera plural y relevante a los atributos de los datos, liberarlos de manera abierta mediante el uso de una licencia para fomentar su reutilización en diversas comunidades y dominios.

Por otra parte, los metadatos son elementos importantes, entre otros, para la descripción y accesibilidad de los datos de investigación. Mayernik (2015) ha estudiado con anterioridad las diferentes configuraciones entre los datos de investigación y los metadatos que se utilizan para describirlos en diferentes tipos de organizaciones y disciplinas inmersas en el contexto académico y de investigación del Reino Unido. Pues mediante el uso de metadatos, se desarrollan perfiles que permiten representar a los atributos de los conjuntos de datos, por este motivo The Data Fair

Principles hacen énfasis en contemplar a los metadatos como parte de su encontrabilidad, accesibilidad, interoperabilidad y reusabilidad.

Además, los metadatos están configurados y son implementados en los sistemas computacionales que son utilizados mediante el uso de programas informáticos, como es el caso de los destinados al desarrollo de repositorios de datos. Pietritch (2017) y Wiley (2014) han abordado la función de los metadatos en el proceso integral de la gestión de los datos de investigación, destacando la importancia de organizar eficientemente a los datos para propiciar su búsqueda y acceso.

A su vez, las políticas para la gestión de los datos de investigación juegan un papel trascendental para fomentar y respaldar el desarrollo de proyectos institucionales para administrar y acceder a los datos que generan sus diferentes comunidades académicas y de investigación.

Al respecto, Leaders Activating Research Networks (LEARN) es un proyecto cuyo objetivo es tomar como base a Roadmap for Research Data, elaborada por la League of European Research Universities (LERU) para construir una infraestructura coordinada en Europa, Latinoamérica y el Caribe con el fin de fomentar la generación de políticas de gestión de datos de investigación (LEARN, 2019, párr. 2). Este proyecto se concentra en la generación de un modelo de política de gestión de datos de investigación y un conjunto de herramientas que soporte la implementación de dicha política.

De la misma manera, DMPTool (2019, párr. 1) es “una aplicación en línea gratuita y de código abierto que ayuda a los investigadores a crear planes de gestión de datos”. Ha sido desarrollada colaborativamente por la National Science Foundation, University of California Curation Center (UC3) en la California Digital Library, DataONE, Digital Curation Centre (DCC-UK), Smithsonian Institution, University of California, Los Angeles Library, University of California, San Diego Libraries, University of Illinois, Urbana-Champaign Library y University of Virginia Library.

Además de eso, desarrollos como la Public Library of Science (PLOS) (<http://plos.org>) y su colección de datos abiertos, ayudan a obtener una visión del alcance de la reutilización y visibilidad de los conjuntos de datos de investigación a través de la web. También, proyectos como DRYAD (<https://datadryad.org/stash>) nos muestran la funcionalidad de los repositorios de datos, al proporcionar la descarga inmediata de los datos de investigación que reúne, la disponibilidad para consultar a las publicaciones que forman parte de dichos datos y consultar las citas que la comunidad le otorga a los conjuntos.

Por todo lo expuesto con anterioridad, este trabajo plantea abordar las siguientes interrogantes:

- ¿Cuáles son los principios que deben considerarse para la descripción de los datos de investigación?
- ¿Cuáles son los elementos del DataCite Metadata Schema que se aplican para la construcción de un perfil descriptivo para datos de investigación?
- ¿Cómo se sistematizan los datos de investigación?

2 OBJETIVOS Y METODOLOGÍA

2.1 Objetivo general

El propósito general de este trabajo es analizar los principios teórico-metodológicos relacionados con la descripción de los datos de investigación.

2.2 Objetivos específicos

Identificar los elementos incluidos en el DataCite Metadata Schema para desarrollar un perfil descriptivo para conjuntos de datos de investigación.

Abordar los elementos que forman parte de la sistematización de los datos de investigación.

2.3 Metodología

Se ha desarrollado una investigación teórica exploratoria apoyada en el análisis documental, para identificar el estado de la cuestión sobre: datos de investigación; citación y descripción de datos de investigación y repositorios de datos. La literatura analizada corresponde a artículos de investigación, estudios de caso y obras monográficas de índole académica. Las fuentes de información que fueron consultadas para la obtención de estos recursos fueron: bases de datos de bibliotecología y estudios de la información, ciencias de la computación e informática; herramientas de

descubrimiento y catálogos en línea de bibliotecas académicas y especializadas; y portales web temáticos. El límite temporal establecido en estas fuentes abarca los años de publicación comprendidos de 2013 a 2019.

A su vez, se identificaron y analizaron los elementos de metadatos para la descripción de conjuntos de datos de investigación que se incluyen en el DataCite Metadata Schema, con el fin de crear una propuesta de perfil descriptivo para conjuntos de datos de investigación, susceptible de aplicarse en la arquitectura de los repositorios de datos.

El perfil descriptivo fue utilizado para describir un conjunto de datos en formato rdf correspondiente a datos bibliográficos y de autoridad de la colección de libros del Instituto de Investigaciones Bibliotecológicas y de la Información (IIBI) dependiente de la Universidad Nacional Autónoma de México (UNAM). Esta colección se encuentra liberada bajo la licencia CC BY-NC-SA 4.0 de Creative Commons y consta de 294 títulos a la fecha de creación del conjunto de datos.

3 DATOS DE INVESTIGACIÓN

En el ámbito de la investigación científica y académica, la aparición del concepto de datos de investigación ha sido utilizada para identificar aquellos datos que son el resultado de experimentos, análisis y procesos metodológicos que emanan de las diferentes actividades de investigación científica y académica. Sin embargo, no existe un consenso que permita establecer su definición formal. Esto debido al amplio contexto disciplinario en donde se originan los datos. Para Henderson (2017, p. 2), los DI “son datos que se recopilan, observan, registran o crean para fines de análisis y producir resultados de investigación originales para un estudio”. Los datos de investigación son almacenados y estructurados en conjuntos de datos que son codificados con diferentes formatos. Esto permite que los datos de investigación sean procesados mediante el uso de software y aplicaciones informáticas especializadas. En algunos campos de investigación, los datos que se generan pueden ser liberados sin un formato establecido, es decir, se obtienen datos en crudo (raw data) que necesitan estructurarse para adquirir una organización acorde al contexto en donde fueron generados. Es pertinente no confundir los formatos de datos con los datos de investigación per se. Pues los primeros forman parte de su encapsulado técnico y los segundos son el resultado de un procesamiento fundamentado en modelos, métodos e instrumentos.

Los datos de investigación, a diferencia de otras expresiones y manifestaciones, se recopilan, observan o crean, con fines de análisis para producir resultados de investigación originales. Además, estos “datos juegan un papel importante en la creación, prueba y revisión de hipótesis, teorías y modelos” (Frederick, 2016, p. 1). Por consiguiente, los datos de investigación pueden ser de índole cuantitativa, cualitativa; de carácter factual o no factual; y ser de naturaleza alfanumérica, textual, audiovisual y digital.

Además, la amplia diversidad de los datos de investigación es una característica que responde al ámbito disciplinario en donde se producen. Por lo tanto, existen datos de investigación con diferente temática. De ahí que surja la necesidad de establecer pautas para su organización y descripción acorde a los atributos que conforman su naturaleza.

Por ejemplo, la concepción de los datos en el ámbito de la medicina no será el mismo que en el de la antropología. Existen diferencias entre los datos y el campo disciplinario y pragmático en donde se generan. El procesamiento de datos suele ser tan especializado que, dentro de un mismo campo de estudio, es necesario realizar un análisis de los atributos de los datos para establecer su representación e identificar sus elementos descriptivos, contemplando para ello, la amplia diversidad de sus fuentes y formatos.

DATOS DE INVESTIGACIÓN	
Fuentes de obtención	Formatos
<ul style="list-style-type: none"> • Encuestas y cuestionarios • Experimentos • Notas de laboratorio • Sensores • Publicaciones arbitradas • Bases de datos • Catálogos en línea • Modelos • Algoritmos • Simulaciones • Evaluaciones • Exámenes • Estadísticas 	<ul style="list-style-type: none"> • .csv • .accdb • .cdf • .odf • .json • .txt • .kml • .xls • .rdf • .pdf • .xml

Tabla II. Principales fuentes y formatos de datos de investigación. Fuente: elaboración propia.

Para ilustrar esto, en la Tabla II se muestran las principales fuentes de obtención y algunos de los formatos mediante los cuales pueden codificarse los datos de investigación. En cada fuente pueden obtenerse datos en diferentes formatos, ya sean libres o propietarios. Por lo cual, es necesario utilizar programas y aplicaciones informáticas para leer conjuntos de datos en específico. “Los formatos a utilizar durante la investigación estarán determinados por el equipo de trabajo, en función de las necesidades de la investigación y los estándares de la disciplina, entre otros factores” (CEPAL, 2019, párr. 5).

El formato csv (comma-separated values en español valores separados por comas) es el más utilizado en el contexto de los datos abiertos, debido a su amplia usabilidad y adaptabilidad al momento de estructurar los datos. Sin embargo, “los valores separados por comas, como formato de archivo, se han utilizado durante décadas para intercambiar información de bases de datos entre computadoras y son anteriores a la iniciativa Open Data” (Mitlohner *et al.*, 2016, p. 72). La estructura más común de los conjuntos de datos es la forma tabular, es decir organizar los datos en filas y columnas para facilitar su manejo. Cuando se plantea la idea de hacer accesibles los datos de investigación, es pertinente considerar el formato en el que serán codificados, su estructuración y la forma en que serán entregados al usuario.

También, es conveniente tomar en cuenta que existe una amplia gama de formatos de datos, los cuales son directamente proporcionales al número de aplicaciones y programas informáticos para procesarlos. “Los formatos de archivo deben elegirse para garantizar el intercambio, el acceso a largo plazo y la preservación de sus datos. Elija estándares abiertos y formatos que sean fáciles de reutilizar” (University of Virginia Library, 2019, párr. 9).

De esta manera, los datos de investigación vistos como objeto de estudio y como base para la generación de nuevos procesos investigativos han sido abordados desde diversas perspectivas disciplinarias. De tal modo que la ciencia intensiva en datos es considerada el cuarto paradigma de la ciencia después de los tres paradigmas interrelacionados de la ciencia empírica, teórica y computacional.

El progreso de la investigación científica no puede concebirse sin el factor tecnológico y computacional que se aplica en el desarrollo de sus diferentes procesos. La investigación actual se basa en la exploración y explotación de datos, para ello las infraestructuras de tecnologías de la información y las herramientas de programas computacionales se utilizan para ayudar a los científicos a administrar, procesar, analizar y compartir datos. Por ejemplo, “varias revistas requieren que se presenten los conjuntos de datos de investigación junto con los artículos” (Patel, 2016, p. 227).

Aunado a ello, el progreso de la investigación siempre ha estado ligado a la explotación y descubrimiento de datos, dando valor a la velocidad de su procesamiento y recolección como parte integral de nuevos hallazgos científicos. “Con la ciencia en proceso de volverse computacional y basada en datos, los principales retos tecnológicos incluyen la necesidad de capturar, analizar, modelar y visualizar información científica de mejor manera” (Fay, 2009, p. 3). De ahí que resulte relevante organizar y proporcionar un acceso eficiente a los datos de investigación que son generados como parte de nuevos procesos creativos. Pues el uso de los datos no puede concebirse sin su correcta organización,

descripción y gestión. Para Cox y Verbaan (2018, p. 4) gestionar datos de investigación consiste en “crear, buscar, organizar, almacenar, compartir y conservar datos dentro de cualquier proceso de investigación”. Es decir, conforman un plan de gestión que contemple las etapas necesarias para facilitar el acceso a estos datos.

Además, los conjuntos de datos de investigación pueden considerarse como recursos con atributos que deben ser descritos mediante estructuras de representación, las cuales deben permitir buscarlos y recuperarlos en un sistema determinado. La sistematización de los datos de investigación es un proceso que consiste en organizar a los datos mediante normas y estándares interoperables, los cuales permitan su recuperación y consecuente acceso por parte de la comunidad científica y académica. De ahí que los datos de investigación tengan que ser organizados para fomentar su posible reutilización mediante la conformación de sistemas abiertos.

En este sentido, “los datos de investigación se abren para dos propósitos: proporcionar evidencia de que la investigación se realizó correctamente y proporcionar datos para su reutilización” (Childs *et al.*, 2014, p. 144). Es pertinente señalar que los datos de investigación pueden estar disponibles para estos dos propósitos sin estar abiertos, por ejemplo, con acceso controlado solo para personas designadas, de esta forma el objetivo de reutilización quedaría rebasado y condicionado. En el siguiente apartado abordamos con mayor detalle este aspecto.

4 SISTEMATIZACIÓN DE LOS DATOS DE INVESTIGACIÓN

El acceso a los datos de investigación enfrenta una serie de retos y oportunidades. El reto principal consiste en convencer a las instituciones de investigación de participar en el modelo del acceso abierto, esto mediante su disposición para desarrollar políticas institucionales que fomenten la accesibilidad abierta a los datos de investigación. Los repositorios de datos de investigación son el resultado de la adopción de estas políticas. Y son el resultado principal de la sistematización de los datos.

La accesibilidad abierta a los datos, se presenta como una oportunidad para fomentar la apertura de las investigaciones que son desarrolladas con fondos públicos y de esta manera conformar un ecosistema de investigación caracterizado por dar amplitud a la visibilidad y reutilización de los datos. Este ecosistema deberá fundamentarse en el uso de políticas institucionales, las cuales deberán contemplar las condicionantes económicas, legales y técnicas que impactan en la creación de sistemas que propicien el acceso a los datos de investigación.

Por consiguiente, un repositorio de datos debe organizar, buscar, recuperar y permitir al usuario acceder a conjuntos de datos de investigación de diversa temática o naturaleza. Estos repositorios son construidos mediante la implementación de un programa computacional que es configurado para tal propósito. Dataverse (<https://dataverse.org/>), Samvera (<https://samvera.org/>) y Fedora (<https://duraspace.org/fedora/>), son tres ejemplos de programas que se dedican a la construcción de este tipo de repositorios.

Características	Descripción
<i>Proporcionan información adicional acerca de su servicio</i>	Un repositorio de datos es extensible en sus funciones. Pues los datos que almacena son un medio para proporcionar servicios adicionales, por ejemplo, son utilizados como fuentes estadísticas, referenciales y como un sistema que fomenta la visibilidad de las investigaciones en el ambiente web.
<i>Enfocados a la organización y acceso a conjuntos de datos</i>	Su objetivo es organizar y fomentar el acceso a los datos de investigación que son codificados a través de diferentes formatos. A diferencia de otros sistemas de información, los repositorios de datos están diseñados para reconocer la autoría de los datos mediante mecanismos de citación como el uso de identificadores persistentes.
<i>Énfasis en la recuperación y visualización de los datos</i>	En los repositorios de datos, la recuperación es multifactorial, es decir, contempla los datos que han sido utilizados para describir a los conjuntos de datos, pero también a los datos que contienen dentro de sus estructuras. Además, la visualización de los datos es un aspecto que permite la interacción mediante representaciones gráficas que hacen posible identificar patrones ocultos mediante la recuperación de los datos.
<i>Ofrece diferentes vías para acceder a los datos</i>	El acceso en los repositorios de datos puede ser abierto, restringido y cerrado. Sin embargo, la visibilidad y reutilización de los datos puede fomentarse mediante su acceso abierto, tomando en consideración la debida citación de los mismos.
<i>Se respaldan en licencias y términos de uso de los datos</i>	Las licencias y términos de uso definen el nivel de manipulación y las condiciones que rigen la utilización de los conjuntos de datos. En estas licencias se establecen las condiciones que deben tomarse en cuenta en materia de propiedad intelectual de los datos.
<i>Están fundamentados en una política</i>	Las políticas de gestión de datos de investigación hacen énfasis en la creación de repositorios de datos. En este sentido, las políticas ponen de manifiesto los principios que regulan la toma de decisiones respecto a los factores técnicos, legales y económicos que inciden en la sistematización de los conjuntos de datos.
<i>Utiliza URI, ORCID, DOI, URN, ARK para que los datos sean persistentes, únicos y citables</i>	La citación de los datos y su persistencia en el ambiente web, son factores que propician su disponibilidad y latente reutilización. Para ello, los repositorios de datos utilizan identificadores persistentes de recursos de fomentan la identificación unívoca de los conjuntos de datos. Es decir, no pueden existir dos o más conjuntos de datos con el mismo identificador.

Tabla III. Características principales de los repositorios de datos de investigación. Fuente: elaboración propia, con datos de <<https://www.re3data.org/faq>>.

En la Tabla III, se exponen las principales características de los repositorios de datos de investigación. La accesibilidad permanente a los datos es una característica que llama la atención. Pues un repositorio de estas características puede proporcionar acceso abierto, cerrado o restringido a los datos de investigación que se han sistematizado. Desde este punto de vista, el hacer uso de un repositorio de datos de investigación, debe contemplar la libre accesibilidad a los datos, de lo contrario el objetivo de la sistematización quedará reducido a la disponibilidad de un almacén de datos de uso local y sin posibilidades de fomentar la visibilidad y reutilización de los datos. Al

respecto, DataCite Metadata Schema ha optado por utilizar el Digital Object Identifier (DOI) para fomentar la persistencia de los datos en el ambiente web y para ser implementado en los repositorios de datos.

Además, el acceso abierto a los datos mediante un repositorio pone de manifiesto la posibilidad de descargar el archivo que contiene el conjunto de datos que en él ha sido almacenado. Esta característica permitirá a los investigadores, reutilizar los datos con miras al desarrollo de nuevos descubrimientos. No obstante, se debe contemplar que “las regulaciones, políticas y pautas establecidas por muchos gobiernos, autoridades federales, así como organizaciones de financiación, sociedades académicas y editoriales, imponen requisitos para la gestión de los datos de investigación” (Redkina, 2019, p. 54). Estos requisitos están relacionados con los aspectos técnicos, legales y económicos que influyen en la utilización de los datos.

Por lo tanto, se requieren acuerdos institucionales que permitan establecer mecanismos de entendimiento entre los diversos actores que intervienen en el desarrollo de investigaciones con fondos públicos y que fomenten la generación de políticas en los contextos de la investigación científica y académica. Pues el resultado de estas políticas se verá reflejado en una mayor generación de repositorios de datos, los cuales permitan conformar un ecosistema digital interoperable para reutilizar, visibilizar y comunicar la investigación a través de los datos que son el reflejo de los resultados alcanzados.

5 PERFIL DESCRIPTIVO PARA CONJUNTOS DE DATOS DE INVESTIGACIÓN

La descripción de los datos de investigación es un proceso intelectual fundamentado en principios y normas que tienen el objetivo de fomentar el óptimo acceso e interoperabilidad de los DI para su latente reutilización. Este proceso ha sido considerado como parte de la gestión de los datos de investigación, en inglés Research Data Management (RDM), sin embargo, esta integración ha sido de carácter superficial, sin establecer principios que permitan profundizar en el corpus teórico-pragmático que debe contemplarse cuando se refiere al proceso intelectual de describir a los datos de investigación que son generados en múltiples contextos de investigación científica y académica.

De ahí, que resulte complejo definir un proceso generalizador para describir a la amplia naturaleza de los datos de investigación. Pues su diversidad en cuanto a su estructura y composición son factores que motivan el establecimiento de un modelo universal para la descripción de datos de investigación. De esta manera, la descripción de los datos de investigación es un proceso que contempla las siguientes fases:

a) Reconocimiento de la naturaleza disciplinaria del conjunto de datos. Es pertinente conocer el campo disciplinario en donde los datos han sido generados. Esto permitirá identificar con mayor certeza los atributos generales y específicos que conforman a los datos que han sido agrupados en un conjunto. Por ejemplo, identificar si se tratan de datos estadísticos, bibliográficos, porcentuales, químicos, biológicos, físicos, matemáticos, etc.

b) Identificación de los atributos del conjunto de datos. La estructura, formato, licencias de uso, tipo de datos (numéricos, alfabéticos, algoritmos, alfanuméricos) son atributos esenciales que permitirán una mejor descripción de los conjuntos de datos, además de la plena identificación de sus puntos de acceso.

c) Identificación de los puntos de acceso del conjunto de datos. La asignación de la autoridad o nombre del creador del conjunto de datos es un elemento trascendental para fomentar su acceso. También, es preciso registrar el título y temática del conjunto. Pues estos elementos son los puntos de acceso que fomentarán su búsqueda y recuperación en el contexto de un sistema.

d) Selección de los metadatos para describir a los conjuntos de datos. Los metadatos son estructuras que permiten representar y registrar los atributos que caracterizan a los conjuntos. DataCite Metadata Schema es una propuesta desarrollada deliberadamente para su aplicación en la descripción y representación de conjuntos de datos de investigación. Con esa finalidad, la tarea consiste en seleccionar elementos de metadatos y construir perfiles para representar conjuntos de diversa naturaleza y temática.

e) Recuperación, visualización y accesibilidad de los conjuntos de datos. La descripción de los conjuntos es un proceso que impacta directamente en su recuperación. Pues los datos que se utilizan para describir a sus atributos son los mismos que se utilizan para efectuar su búsqueda en un sistema que propicie su recuperación. Es decir, los atributos de los conjuntos son directamente proporcionales a los datos que utilizamos para recuperarlos, por ejemplo, mediante el uso de palabras y números, realizamos búsquedas que nos arrojen resultados similares y en concordancia con el autor, título y tema de un conjunto en específico. Por lo tanto, se estima que una correcta descripción de los conjuntos fomentará su óptima recuperación. A su vez, la visualización de los datos de investigación es un gran reto para los actuales repositorios de datos, pues si hablamos de una efectiva reutilización de estos datos, es necesario que los sistemas que los almacenan también ofrezcan mecanismos para su visualización gráfica. De acuerdo con Ferster

(2013, p. 5), la visualización “implica la reducción de información en bruto, como datos, en elementos gráficos más simples que usan variables espaciales, como posición, tamaño, forma y color, para revelar relaciones visuales y patrones implícitos pero ocultos dentro de los datos”. De esta manera, la accesibilidad a los datos de investigación debe contemplar su reutilización mediante el uso de licencias abiertas que propicien su libre acceso y promuevan la eliminación de barreras económicas, legales y técnicas que interfieran en su acceso. Esto significa un gran reto y deja entrever un ambicioso cambio de paradigma en los sistemas para la publicación y comunicación de la ciencia. De ahí que, el cuarto paradigma de la ciencia se encuentre fundamentado en el descubrimiento intensivo de datos, pues “la meta es tener un mundo en el que toda la literatura científica esté en línea, todos los datos científicos estén en línea, y que interactúen” (Hey, Tansley y Tolle, p. xxxii).

Las fases descritas previamente permiten establecer la generación de perfiles de metadatos que sirvan para describir a los conjuntos de datos de investigación. El esquema de metadatos de DataCite en su versión núcleo 4.1 (DataCite Metadata Working Group, p. 8), consta de 19 propiedades de metadatos, divididas en tres diferentes niveles de uso: M=Mandatory=Obligatorio; R=Recommended=Recomendado; O=Optional=Opcional.

ID	PROPIEDAD	USO
1	Identificador	Obligatorio
2	Creador	Obligatorio
3	Título	Obligatorio
4	Editor	Obligatorio
5	Año de publicación	Obligatorio
6	Tema	Recomendado
7	Colaborador	Recomendado
8	Fecha	Recomendado
9	Idioma	Opcional
10	Tipo de recurso	Obligatorio
11	Identificador alternativo	Opcional
12	Identificador relacionado	Recomendado
13	Tamaño	Opcional
14	Formato	Opcional
15	Versión	Opcional
16	Derechos	Opcional
17	Descripción	Recomendado
18	Geolocalización	Recomendado
19	Referencia de financiamiento	Opcional

Tabla IV. Propiedades de metadatos del esquema DataCite. Fuente: elaboración propia, con datos obtenidos de DataCite Metadata Schema Documentation, versión 4.1. Disponible en: <http://schema.datacite.org/meta/kernel-4.1/doc/DataCite-MetadataKernel_v4.1.pdf>.

Cada propiedad de metadato tiene subpropiedades que permiten especificar con mayor detalle la descripción de los conjuntos de datos de investigación. Las subpropiedades son controladas mediante el principio de ocurrencia (Occ), el cual indica la cardinalidad y cantidad de veces que puede repetirse una subpropiedad. Estos elementos son expresados de la siguiente manera: 0-n= opcional y repetible, 0-1= opcional, pero no repetible, 1-n=requerido y repetible, 1= requerido, pero no repetible.

El esquema de DataCite utiliza eXtensible Markup Language (XML) para etiquetar y construir la sintaxis de las descripciones de los conjuntos. Esto fomenta el intercambio y adaptación de las descripciones en un contexto sistematizado. En un nivel orientado a la sistematización y recuperación de los datos, el etiquetado de las descripciones es sumamente significativo pues se fomenta su vinculación y acceso en el entorno web, mediante una arquitectura interoperable. Además, la citación de los datos de investigación es otro aspecto en el que DataCite pone especial énfasis, pues esto fomenta el reconocimiento intelectual y la reutilización de los datos en diversos dominios disciplinarios.

Todo	Autor	Título
☆	51. Rodríguez García, Ariel Alejandro	Tendencias multidisciplinares del uso de los metadatos
☆	52. Martínez Arellano, Filiberto Felipe	Tendencias en la teoría y práctica de la catalogación bibliográfica: memoria del Segundo Encuentro Internacional de Catalogación, 12 al 14 de septiem
☆	53. Ramírez Leyva, Elsa Margarita	Tendencias de la lectura en la universidad
☆	54. Martínez Arellano, Filiberto Felipe	Tendencias de la investigación en bibliotecología y documentación en México y España: memoria del Tercer Seminario Hispano-Mexicano de investiga
☆	55. Martínez Arellano, Filiberto Felipe	Tendencias de la investigación en bibliotecología y documentación en México y España: memoria del 5º Seminario Hispano-mexicano de Investigación
☆	56. Hernández Salazar, Patricia	Tendencias de la alfabetización informativa en Iberoamérica
☆	57. Rodríguez Gallardo, Adolfo	Tecnologías de la información y brecha digital en México 2001-2005
☆	58. Torres Vargas, Georgina Araceli	Software libre: miradas desde la bibliotecología y estudios de la información
☆	59. Martínez Arellano, Filiberto Felipe	Sociedad, información y bibliotecas: memoria del XXII Coloquio de Investigación Bibliotecológica y de la Información, 22-24 de septiembre de 2004
☆	60. Hernández Salazar, Patricia	Significados e interpretaciones de la información desde el usuario
☆	61. Martínez Arellano, Filiberto Felipe	Servicios bibliotecarios en América Latina: tres casos prominentes
☆	62. Ramírez Leyva, Elsa Margarita	Seminario Lectura: Pasado, Presente y Futuro, 29 de septiembre al 2 de octubre de 2003: memoria
☆	63. Martínez Arellano, Filiberto Felipe	Seminario INFOBILA como Apoyo a la Investigación y Educación Bibliotecológica en América Latina y el Caribe, 16, 17 y 18 de marzo de 2005: memoria
☆	64. García Aguilar, María Idalia	Secretos del estante: elementos para la descripción bibliográfica del libro antiguo
☆	65. Calva González, Juan José	Satisfacción de usuarios: la investigación sobre las necesidades de información
☆	66. Morales Campos, Estela Mercedes	Rescate y difusión de la diversidad cultural en el mundo global
☆	67. Campos, Morales	Regulaciones que impactan la infodiversidad y el acceso abierto a la información en la sociedad global y multicultural
☆	68. Ríos Ortega, Jaime	Procesos revolucionarios, bibliotecas y movimientos culturales
☆	69. Rodríguez Reséndiz, Perla Olivia	Preservación digital sustentable de archivos sonoros
☆	70. Voutsás Márquez, Juan	Preservación del patrimonio documental digital en México
☆	71. Rodríguez Reséndiz, Perla Olivia	Preservación de documentos sonoros y audiovisuales de origen digital
☆	72. Gorbea Portal, Salvador	Potencialidades de investigación y docencia iberoamericanas en ciencias bibliotecológica y de la información
☆	73. Morales Campos, Estela Mercedes	La posverdad y las noticias falsas : el uso ético de la información
☆	74. Sánchez Vanderkast, Egbert John	Políticas de información en las universidades públicas estatales
☆	75. Ríos Ortega, Jaime	Oportunidades y retos en la formación, investigación y aplicación del conocimiento bibliotecológico

Figura 1. Estructura básica del conjunto de datos de libros del IIBI. Fuente: elaboración propia.

En la Figura 1 puede observarse la estructura básica del conjunto de datos que fue descrito con el perfil diseñado. El conjunto cuenta con filas y columnas que contienen datos estructurados referentes a los autores y títulos de los libros del IIBI.

A su vez, en la Figura 2 se muestra el formato rdf del mismo conjunto, el cual fue construido mediante el vocabulario Bibframe (<https://lov.linkeddata.es/dataset/lov/vocabs/bf>). Este conjunto de datos fue desarrollado con la intención de establecer vinculaciones de significado entre los datos de autoridad y bibliográficos pertenecientes a la colección de libros del IIBI.

```

<?xml version="1.0" encoding="UTF-8"?><rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:owl="http://www.w3.org/2002/07/owl#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:bibframe="http://id.loc.gov/ontologies/bibframe/"
xmlns:foaf="http://xmlns.com/foaf/0.1/"><rdf:Description
rdf:about="http://127.0.0.1:3333/lafuente-lpez-ramiro">
<rdf:type
rdf:resource="http://id.loc.gov/ontologies/bibframe/Person"/>
<bibframe:title rdf:resource="http://127.0.0.1:3333/los-
sistemas-bibliotecologicos-de-clasificacion"/>
</rdf:Description><rdf:Description
rdf:about="http://127.0.0.1:3333/iibi-unam"> <rdf:type
rdf:resource="http://id.loc.gov/ontologies/bibframe/Person"/>
<bibframe:title
rdf:resource="http://127.0.0.1:3333/vocabulario-controlado-en-
bibliotecologia-ciencia-de-la-informacion-y-temas-afines"/>
</rdf:Description><rdf:Description
rdf:about="http://127.0.0.1:3333/barquet-tellez-concepcion">
<bibframe:title rdf:resource="http://127.0.0.1:3333/catalogo-
de-tesis"/> <rdf:type
rdf:resource="http://id.loc.gov/ontologies/bibframe/Person"/>
</rdf:Description><rdf:Description
rdf:about="http://127.0.0.1:3333/shera-jesse-hauk">
<bibframe:title rdf:resource="http://127.0.0.1:3333/los-
fundamentos-de-la-educacion-bibliotecologica"/> <rdf:type

```

Figura 2. Sintaxis rdf de la estructura del conjunto de dato de libros del IIBI. Fuente: elaboración propia.

En la Tabla V, puede observarse el resultado de la descripción de un conjunto de datos bibliográficos y de autoridad, relacionados con la colección de libros digitales del IIBI. Estos datos reflejan el contenido intelectual de las investigaciones desarrolladas en esta entidad académica. De esta manera, se seleccionaron del esquema de metadatos de DataCite, solo aquellas propiedades y subpropiedades a considerar durante el proceso descripción de los atributos de este conjunto de datos.

ID	PROPIEDADES Y SUBPROPIEDADES	Occ	DESCRIPCIÓN
1	Identificador	1	https://data_repository/iibi_digital_books
1.1	Tipo de identificador	1	URI
2.1	Nombre del creador	1	Ávila, Eder
2.1.1	Tipo de nombre	0-1	Personal
2.2	Nombre de pila	0-1	Eder
2.3	Apellido	0-1	Ávila
2.4	Identificador de nombre	0-n	https://orcid.org/0000-0001-5439-0495
2.4.1	Esquema del identificador de nombre	1	ORCID
2.4.2	URI del esquema	0-1	https://orcid.org/
2.5	Afiliación	0-n	Instituto de investigaciones Bibliotecológicas y de la Información
3	Título		Libros digitales IIBI-UNAM
4	Editor	1	Universidad Nacional Autónoma de México
4	Editor	1	Instituto de Investigaciones Bibliotecológicas y de la Información
5	Año de publicación	1	2019
6	Tema	0-n	Bibliotecología
6	Tema	0-n	Estudios de la Información
8	Fecha	0-n	2019-10-17
8.1	Tipo de fecha	1	Creado
9	Idioma	0-1	spa
10	Tipo de recurso	1	Conjunto de datos
10.1	Tipo general de recurso	1	Text
13	Tamaño	0-n	84.9 kb
14	Formato	0-n	rdf
16	Derechos	0-n	CC BY-NC-SA 4.0
16.1	URI de los derechos	0-1	https://creativecommons.org/licenses/by-nc-sa/4.0/
17	Descripción	0-n	Conjunto de datos estructurado con el vocabulario Bibframe 2.0 y codificado en rdf. Contiene los autores y respectivos títulos de la colección de libros digitales del Instituto de Investigaciones Bibliotecológicas y de la Información (IIBI) de la Universidad Nacional Autónoma de México (UNAM).
17.	Tipo de descripción	1	Resumen

Tabla V. Descripción de un conjunto de datos rdf, perteneciente a la colección de libros digitales del IIBI. Fuente: elaboración propia.

La mayoría de los atributos fueron identificados plenamente mediante el análisis de la estructura de los datos que contiene el archivo rdf mediante el cual estaba codificado el conjunto de datos. Este proceso requirió utilizar programas informáticos especializados en el uso y manipulación de datos rdf, pues la descripción del conjunto de datos es un procedimiento que debe remitirse al análisis del archivo digital en sí mismo.

La propiedad de derechos y subpropiedad URI de los derechos, requieren especial atención. Pues para registrar una correcta descripción de los conjuntos de datos y fomentar su reutilización y acceso abierto, es necesario definir una política de libre acceso a los datos que contemple, además, el uso de la licencia que va a respaldar su libre accesibilidad.

Resulta pertinente tener en cuenta que publicar y liberar datos para su latente reutilización, es un asunto más complejo que poner a disposición abierta enormes cantidades de publicaciones académicas y arbitradas. Pues los conjuntos de datos requieren de una mayor confiabilidad e integridad al momento de liberarse en el ambiente de la web para ello es necesario optar por un esquema flexible que permita citar y dar reconocimiento a los creadores primigenios de los datos de investigación. Al respecto, Starr y Gastl (2011, párr. 9) manifiestan que uno de los principales objetivos del DataCite Metadata Schema consiste en “recomendar un formato de cita estándar para conjuntos de datos, basado en un pequeño número de propiedades requeridas para el registro del identificador”. Como se ha mencionado con anterioridad, el identificador utilizado por este esquema es el DOI.

Aunado a ello, la reutilización de los conjuntos de datos requiere tener acceso a metadatos apropiados y a la información contextual para que los datos puedan ser comprendidos por los demás investigadores. Pues los datos son fuente para la obtención de nuevos resultados. De ahí que su descripción sea un proceso que debe fundamentarse en principios, normas y estándares internacionales para asegurar su interoperabilidad y visibilidad.

6 CONCLUSIONES

Los datos de investigación son el resultado de procesos rigurosos relacionados con el análisis de fenómenos, el desarrollo de experimentos y la aplicación de metodologías. Se encuentran almacenados y codificados en conjuntos de datos que pueden estar estructurados o no estructurados. La temática y naturaleza de los datos de investigación pueden ser muy diversas, de ahí que existan datos científicos, datos estadísticos o datos que remitan a recursos de información documental, o bien, datos que formen parte de un proceso de investigación relacionado con una disciplina en específico.

Los conjuntos de datos de investigación tienen atributos de autoridad, identificativos y temáticos que tienen que ser descritos para fomentar su recuperación y accesibilidad en un entorno digital sistematizado. En este sentido, la descripción de los datos de investigación es un proceso intelectual y normalizado que tiene el propósito de identificar y representar a los puntos de acceso (autor, título y tema) de los conjuntos para propiciar su uso y acceso.

El acceso a los datos de investigación puede ser abierto, restringido y cerrado. Sin embargo, para fomentar su reutilización, es necesario que las instituciones académicas y de investigación contribuyan a la generación de políticas de acceso abierto a sus datos de investigación, pues de esta manera se propiciará la creación de repositorios de datos que permitan recuperar, acceder y utilizar conjuntos de datos para su latente aprovechamiento.

Bajo esta premisa, la citación de los conjuntos de datos de investigación es un aspecto que permitirá dar mayor visibilidad la investigación realizada, dar el debido crédito a los investigadores y creadores de los datos y seguir el impacto de un determinado conjunto de datos mediante métricas que permitan evaluar los resultados de su reutilización.

Se estima que, en un futuro cercano citar los datos de investigación permitirá obtener posibles incentivos por su utilización tras su publicación, tal y como ocurre con otros productos de investigación científica y académica. “Las normas para citar datos están tomando forma para mejorar el intercambio y el reconocimiento. Dado que la atribución y el impacto son elementos vitales de la Academia, se necesita fomentar constante la cita de los datos. Al hacerlo, se puede acreditar a un investigador no solo por su investigación directa, sino también por la reutilización de los datos por parte de otra persona” (Levine, 2013, p. 135).

Esto parece que pasará por la obligada configuración de un nuevo ecosistema de publicación y comunicación del conocimiento, el cual tendrá que enfrentar desafíos relacionados con los aspectos legales, económicos y técnicos que influyen en la libre accesibilidad a los datos de investigación. La gestión de los datos de investigación está suponiendo un reto a corto plazo para las universidades e institutos de investigación. La formulación de políticas para la creación de repositorios de datos de investigación será una propuesta para enfrentar los desafíos de la descripción y accesibilidad abierta a este tipo de datos.

7 BIBLIOGRAFÍA

- CALLAGHAN, S. Processes and procedures for data publication: a case study in the Geosciences. *The International Journal of Digital Curation*, 2013, vol. 8, nº 1, p. 193-203.
- CEPAL. Biblioguias. Gestión de datos de investigación: formatos abiertos y cerrados [en línea]. Disponible en: <<https://biblioguias.cepal.org/gestion-de-datos-de-investigacion/formatos>> [Consulta: 9 de octubre de 2019]
- CHILDS, S. *et al.* Opening Research data: issues and opportunities. *Records management journal*, 2014, vol. 24, nº 2, p. 142-162.
- COX, A. y VERBAAN, E. *Exploring research data management*. Londres: Facet Publishing, 2018.
- CSUC (Consortio de Servicios de la Universidad de Cataluña). FAIR X FAIR: requisitos factibles, alcanzables e implementables para un repositorio de datos de investigación FAIR [en línea]. Disponible en: <https://recercat.cat/bitstream/handle/2072/356460/InformeFxF_maquetada_ESP.pdf?sequence=5> [Consulta: 23 de octubre de 2019]
- DMPTool. About [en línea]. Disponible en: <https://dmptool.org/about_us> [Consulta: 24 de octubre de 2019]

- FAY, D. Introducción. En HEY, T.; TANSLEY, S. y TOLLE, K. (eds.). *El cuarto paradigma: descubrimiento científico intensivo en datos*. Estados Unidos de América: Microsoft Research, 2009, p. 3-5.
- FERSTER, B. *Interactive visualization: insight through inquiry*. London: MIT Press, 2013.
- FREDERICK, D. Data, Open Science and Libraries. *Library Hi-Tech News*, 2016, vol. 33, nº 8, p. 11-16.
- HARARI, Y. *Homo Deus*. Barcelona: Penguin Random House, 2016.
- HENDERSON, M. *Data management: a practical guide for librarians*. Maryland: Rowman & Littlefield, 2017.
- HEY, T.; TANSLEY, S. y TOLLE, K. Jim Gray sobre la e-ciencia: un método científico transformado. En HEY, T.; TANSLEY, S. y TOLLE, K. (eds.). *El cuarto paradigma: descubrimiento científico intensivo en datos*. Estados Unidos de América: Microsoft Research, 2009, p. xvii-xxxiii.
- LEADERS ACTIVATING RESEARCH NETWORKS. Project Summary [en línea]. Disponible en: <<http://learnrdm.eu/en/about/project-summary/>> [Consulta: 24 de octubre de 2019]
- LEVINE, M. Copyright, open data and availability-usability gap. En RAY, J. (ed.). *Research Data Management*. Indiana: Purdue University Press, 2013, p. 129-147.
- MAYERNIK, M. Research Data and Metadata Curation as Institutional Issues. *Journal of Association for Information Science and Technologies*, 2015, vol. 67, nº 4, p. 973-993.
- MITLOHNER, J.; NEUMAIER, S.; UMBRICH, J. y POLLERES, A. Characteristics of Open Data CSV Files. 2nd International Conference on Open and Big Data (OBD), 2016, p. 72-79.
- PATEL, D. Research data management: a conceptual framework. *Library Review*, 2016, vol. 65, nº 4/5, p. 226-241.
- PETRITSCH, B. Metadata for Research: Data in Practice. *VOEB: Mitteilungen*, 2017, vol. 10, nº 2, p. 200-207.
- R3DATA. Frequently Asked Questions [en línea]. Disponible en: <<https://www.re3data.org/faq>> [Consulta: 22 de octubre de 2019]
- REDKINA, N. Current trends in research data management. *Scientific and technical Information processing*, 2019, vol. 46, nº 2, p. 53-58.
- STARR, J. y GASTL, A. isCitedBy: A Metadata Scheme for DataCite. *D-Lib Magazine*, 2011, vol. 17, nº 1/2 [en línea]. Disponible en: <<http://www.dlib.org/dlib/january11/starr/01starr.html>> [Consulta: 3 de diciembre de 2019]
- UNIVERSITY OF VIRGINIA LIBRARY. Research data services + sciences. Data types and files formats [en línea]. Disponible en: <<https://data.library.virginia.edu/data-management/plan/format-types/>> [Consulta: 9 de octubre de 2019]
- WILEY, C. Metadata Use in Research Data Management". *Bulletin of the Association for Information Science and Technology*, 2014, vol. 40, nº 6, p. 38-40.
- WILKINSON, M. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 2016, vol. 3, nº 1, p. 1-9.