
ESTUDIOS / RESEARCH STUDIES

Aspectos metodológicos de los datos abiertos de investigación: análisis de los conjuntos de datos de la colección SciELO incluidos en Figshare

Críspulo Travieso Rodríguez*, Ronaldo Ferreira Araújo**

*Dpto. Biblioteconomía y Documentación, Universidad de Salamanca
Correo-e: ctravieso@usal.es | ORCID iD: <https://orcid.org/0000-0002-0774-0728>

**Instituto de Ciências Humanas, Comunicação e Artes (ICHCA), Universidad Federal de Alagoas
Correo-e: ronaldfa@gmail.com | ORCID iD: <https://orcid.org/0000-0003-0778-9561>

Recibido: 19-06-2018; 2ª versión: 13-11-2018; Aceptado: 08-01-2019.

Cómo citar este artículo/Citation: Travieso Rodríguez, C.; Araújo, R. F. (2019). Aspectos metodológicos de los datos abiertos de investigación: análisis de los conjuntos de datos de la colección SciELO incluidos en Figshare. *Revista Española de Documentación Científica*, 42 (3), e242. <https://doi.org/10.3989/redc.2019.3.1597>

Resumen: Los datos abiertos de investigación constituyen una oportunidad para el desarrollo colaborativo de la ciencia y abren nuevas perspectivas para la validación y difusión de los resultados científicos. La reutilización efectiva de estos datos requiere conocer tanto los procedimientos empleados para su obtención como las métricas de uso y consumo que van acumulando. El presente trabajo toma la colección íntegra de datos abiertos (348 datasets) de las revistas de SciELO disponibles para su consulta desde Figshare, con el objetivo de examinar sus metodologías de investigación y la naturaleza de los datos, en función de cómo han sido obtenidos y según sus categorías temáticas. Asimismo, se recopilan sus indicadores de uso y citación, los formatos de archivo y sus licencias de utilización. Los resultados muestran una mayoría de estudios cuantitativos frente a los cualitativos. Los conjuntos de datos más frecuentes son los obtenidos mediante observación, seguidos de los experimentales y derivados, aunque con variación en función de las disciplinas científicas, mientras que los indicadores de uso resultaron ser poco significativos aún para toda la colección.

Palabras clave: datos de investigación; datos abiertos; métodos de investigación; descripción de datos; datasets; uso de datos; repositorios de datos.

Methodological issues of open research data: analysis of the datasets from SciELO included in Figshare

Abstract: Open research data represent an opportunity to share scientific knowledge and to provide new perspectives for validation and dissemination of scientific results. Effective re-use of these data implies to know both the research methods applied to obtain them and the measures of usage and citation that are systematically gathered. Based on the whole catalogue of open research data (348 datasets) from SciELO journals available in Figshare, this study aims to explore their research methods and the types of data according to data collection techniques and subject categories. It also analyses usage and bibliometric indicators, file formats and dataset licensing. Results show that quantitative data are more frequent than qualitative data. The most usual data were those obtained by observation techniques, followed by experimental and derived data. This figures varied according to the different scientific areas, but usage indicators came out not to be significant for the whole set of data files.

Keywords: research data; open data; research methods; data description; datasets; data usage; data repositories.

Copyright: © 2019 CSIC. Este es un artículo de acceso abierto distribuido bajo los términos de la licencia de uso y distribución Creative Commons Reconocimiento 4.0 Internacional (CC BY 4.0).

1. INTRODUCCIÓN

La difusión en abierto y el uso compartido de los datos de investigación se enmarcan en un contexto de cambio de la propia concepción de la comunicación científica. Una prueba más de este escenario es el reciente acuerdo al que han llegado el repositorio *Figshare* y la biblioteca digital SciELO, según el cual se da acceso en abierto, desde principios de 2018, a los datasets de las revistas que esta última plataforma recoge.

Este tipo de alianzas y acciones de colaboración se explican en gran medida por la extensión del concepto de ciencia abierta (*open science*), fenómeno que engloba tanto las nuevas formas de transmitir conocimiento como las iniciativas políticas e institucionales que las impulsan. Su fin último es conseguir que la investigación científica sea más eficiente, transparente e interdisciplinar, potenciando así su impacto social (Ramjoué, 2015), y atañe a todos los agentes implicados en su desarrollo (gobiernos y administraciones, comunidad científica, agencias y centros de investigación, editores).

Entre las propuestas para concretar una definición de ciencia abierta destaca por su carácter integrador la que ofrecen Vicente-Sáez y Martínez-Fuentes (2018), que, tras una revisión sistemática de 75 artículos e informes previos, concluyen definiéndola como "conocimiento transparente y accesible que se comparte y desarrolla a través de redes de colaboración". Varias de las referencias empleadas por estos autores provienen de la Comisión Europea, que constituye uno de los órganos políticos que de manera más decidida ha estimulado la adopción de prácticas de ciencia abierta, como demuestran los requisitos del programa Horizon 2020 en cuanto a publicación de los resultados en acceso abierto y la obligación de que los datos de investigación derivados estén disponibles para ser reutilizados (European Commission, 2016; 2016b).

Con similar intención de articular los distintos enfoques y elementos intervinientes, Fecher y Friesike (2014) proponían un modelo que distingue cinco escuelas de pensamiento para abordar teóricamente el concepto *open science*. Para ello parten de las siguientes perspectivas: pragmática, pública, democrática, de infraestructuras y de evaluación. Cada una de ellas pone el acento en algunos de los objetivos perseguidos con la implementación de políticas de ciencia abierta. La disponibilidad de los datos de investigación en abierto aparece reflejada como una pieza sustancial tanto en la perspectiva democrática – que defiende el acceso libre y la distribución igualitaria del conocimiento científico –, como en la pragmática – que está centrada en la creación del conocimiento de forma colaborativa para que sea más eficiente .

Dentro de las posibles líneas de análisis en este ámbito, el objetivo principal de este trabajo es conocer y estudiar la tipología, origen y características de los datos abiertos de investigación contenidos en la colección de SciELO y disponibles desde *Figshare*. De manera secundaria, también se abordan aspectos relacionados con su uso, como forma de calibrar su repercusión y fomentar su empleo.

2. MARCO TEÓRICO

Datos abiertos de investigación: aproximaciones teóricas e iniciativas

Los datos abiertos de investigación heredan las características de los *open data* en cuanto a la garantía de transparencia y a su relación con la rendición de cuentas y la reutilización (Peset y Fernández-López, 2014), especialmente al hablar de investigaciones llevadas a cabo mediante fondos públicos. Pero, además, se definen por el contexto en el que surgen y al que sirven. Así, pueden considerarse datos de investigación "todo aquel material que ha sido registrado durante la investigación, reconocido por la comunidad científica y que sirve para certificar los resultados de la investigación que se realiza" (Torres-Salinas y otros, 2012). Si a ello sumamos su carácter abierto, obtendremos que han de poder ser utilizados, reutilizados y redistribuidos libremente por cualquier persona, siempre que se reconozca su autoría y que se vuelvan a compartir una vez se hayan transformado durante un uso posterior. En palabras de Melero y Hernández-San Miguel (2014), el acceso a estos datos evita duplicar esfuerzos en su obtención, y su agregación o transformación puede generar nuevos datos y favorecer otras interpretaciones y colaboraciones interdisciplinares. En todo caso, será la descripción que se asocia y acompaña a esos conjuntos de datos abiertos la que permitirá que estos puedan ser recuperados, reutilizados; consecuentemente, esta puede ser considerada en sí misma indicio de la calidad de un dato de investigación abierto (Peset Mancebo y otros, 2017).

Desde distintas esferas se ha señalado la importancia de una adecuada gestión de estos datos, poniendo de manifiesto la necesidad de garantizar los medios y los procedimientos para su preservación y reutilización (Estevão y otros, 2017). Además del programa de investigación citado, la agenda política comunitaria ha continuado incidiendo en esa línea a través de sucesivos informes y publicaciones, que se han materializado, entre otras acciones, en proyectos como la plataforma FOSTER Plus, el proyecto *Open Research Data* (ORD) sobre acceso y uso de datos de investigación o las recomendaciones planteadas por *Open Science Policy Platform* (OSPP).

Al margen de las iniciativas europeas, la *National Science Foundation* estadounidense exige que sus proyectos de investigación vayan acompañados de un plan de gestión de datos que permita el acceso a los mismos. En 2016 un grupo de agencias de investigación anunciaban el lanzamiento de *Open Research Funders Group* (ORFG), un consorcio cuyo objetivo era aumentar el acceso abierto a los resultados y datos de investigación, de forma que sean compartidos y reproducibles para incentivar la innovación y el progreso social. En una dirección similar apuntan sendos informes de organismos tan relevantes como Naciones Unidas (UN Data Revolution, 2014) y la OCDE (2015).

Mención aparte merecen los Principios FAIR sobre buenas prácticas en la gestión de datos científicos, que se corresponden con las siglas en inglés de localizables, accesibles, interoperables y reutilizables; estos principios están siendo incorporados al funcionamiento de la plataforma *European Open Science Cloud* (EOSC). Igualmente, el proyecto europeo LEARN toma como punto de partida los postulados marcados por la LERU (*League of European Research Universities*) para construir una estructura coordinada para datos de investigación en Europa, Latinoamérica y Caribe, desarrollando una política a tal efecto y un conjunto de herramientas para su aplicación (Ayris, 2016).

En lo que se refiere a la forma de hacer públicos esos datos se cuenta con varias vías. En primer lugar, existen los repositorios destinados específicamente a datos de investigación, que pueden ser institucionales o gestionados por intermediarios, adscritos a una circunscripción territorial, de temática generalista o especializados en una materia concreta. Varios trabajos analizan las características de este tipo de repositorios, ofreciendo una visión cuantitativa en función de distintos criterios, como puedan ser su tipología, materia, condiciones de acceso y licencias, inclusión en directorios (como *re3data – Registry of Research Data Repositories*), servicios ofrecidos, software e interfaces de búsqueda, etc. (Kindling y otros, 2017; Assante y otros, 2016; Beagrie y Houthton, 2014; Marcial y Hemminger, 2010).

En segundo lugar, otra forma de hacer públicos estos datos es incluirlos en las páginas de las propias revistas científicas. Los editores pueden establecer requisitos a los autores para que ofrezcan los archivos de datos sobre los que sustentan las conclusiones de sus artículos. Ejemplo de ello tenemos en el grupo de revistas incluidas en *Dryad, Journal of the Royal Society, Interface, Nature, PloS, Science and National Academies Press* (Victoria University, 2015). Por último, también se publican revistas, llamadas *data journals*, en las que la

información suministrada se corresponde con conjuntos de datos de investigación que se acompañan de información adicional para su reutilización por parte de la comunidad científica (García-García y otros, 2015; Candela y otros, 2015).

Tipologías de datos abiertos de investigación

Los datos de investigación pueden tomar varias formas y ser categorizados en función de distintos criterios (Borgman, 2012). Estos aspectos pueden contemplar, entre otras, cuestiones relacionadas con la metodología adoptada para su obtención, su presentación y formatos de archivo, la naturaleza de los datos, el campo del conocimiento al que pertenecen o sus restricciones de acceso y reutilización. Del mismo modo en que tradicionalmente se establecen estas condiciones para poder comparar bases de datos y artículos científicos desde el punto de vista de la evaluación bibliométrica, también es conveniente tener en cuenta esta gran diversidad de aproximaciones antes de cotejar el alcance de estos datos y comparar posteriormente su incidencia e impacto de forma consistente.

Además de la distinción clásica de la metodología científica entre datos cualitativos y cuantitativos (Ríos-Hilario y Travieso-Rodríguez, 2013), y dada la riqueza metodológica posible en los conjuntos de datos de investigación, puede establecerse otra tipificación atendiendo, en esta ocasión, a la técnica empleada para obtener los datos. De este modo, es posible diferenciar entre las siguientes cinco instancias (Stony Brook University, 2018; Krier y Straser, 2014; Melero y Hernández San Miguel, 2014; FECYT, 2012):

- Datos observacionales. Se recogen *in situ*, no pueden ser recreados, duplicados o vueltos a recopilar en las mismas condiciones. Algunos ejemplos serían los resultados de encuestas, las observaciones sensoriales realizadas por el investigador o con dispositivos que miden temperaturas, sonidos, etc. o transcripciones de entrevistas y grupos de discusión.
- Datos experimentales. Se recogen bajo circunstancias controladas durante el desarrollo de un experimento o en condiciones de laboratorio. En la mayoría de los casos han de ser reproducibles, si bien con un coste elevado y con la condición habitual de requerir instrumentos especializados. Las secuencias genéticas, cromatogramas o espectroscopias podrían ser ejemplos de este tipo de datos.
- Datos derivados o compilados. Suelen incluir datos de entrada, softwares o programas concretos y sus resultados. Ello hace que sean

reproducibles si se cuenta con el soporte tecnológico adecuado. Es el caso de la aplicación de técnicas de minería de datos, tratamiento de corpus textuales, extracciones de bases de datos o modelos 3D.

- Datos simulados. Son resultado de usar un modelo para estudiar el comportamiento de un sistema real o teórico. En este tipo de datos los datos de entrada pueden ser más significativos que los propios datos resultantes. Como ejemplo de ellos se citan los modelos climáticos, los modelos económicos o los modelos geoquímicos.
- Datos referenciados. Son conjuntos de datos validados que ya han sido publicados o conservados. Entre estos se incluyen los bancos de datos genéticos, estructuras químicas, censos, estadísticas oficiales publicables y portales de datos abiertos.

Desde un punto de vista menos conceptual, otro aspecto fundamental es el referido a los formatos de archivo en que se comparten estos datos, dado que en ocasiones estos estándares determinan la utilidad o reutilización efectiva de los mismos y las posibilidades de importar o exportar esa información codificada. En ese sentido, el abanico de posibilidades es muy amplio (en la Tabla I se ofrece un listado como muestra de esta circunstancia), dando lugar a distintas clasificaciones. En cualquier caso, lo habitual es que los repositorios de datos no impongan restricciones en cuanto a formatos de archivo a la hora de su almacenamiento.

Otro modo de sistematizar el análisis de los datos abiertos de investigación es el propuesto por la base de datos *Data Citation Index* (Thomson Reuters, 2012), que distingue, además de los repositorios, dos tipos documentales de registros

de datos: *dataset* o conjunto de datos, y *data study* o informe de datos. Si bien los primeros son los que directamente se identifican con un fichero de datos, que puede estar o no vinculado a un artículo al que acompañan e ilustran o fundamentan, los informes de datos comprenden mayor volumen de información sobre el contexto y la forma de extracción de los datos, los autores, referencias bibliográficas, cobertura geográfica y temporal o referencias. En definitiva, son objetos digitales que incluyen una descripción exhaustiva de los propios datos abiertos. En la práctica, cualquier fichero de datos es susceptible de ser citado si está identificado de forma unívoca y coherentemente descrito, lo que abre la puerta a establecer métricas sobre su uso y citación.

Descripción de datos abiertos y métricas de uso

La disponibilidad de los datos abiertos debe ir acompañada de una descripción documental de sus características. Esto permitirá no solo contextualizar y validar los datos ofrecidos sino resolver eficazmente su recuperación posterior y eventual reutilización. Generalmente, la documentación que acompaña a los datasets viene determinada por los metadatos que se requieren en las distintas plataformas para su depósito. En 2016 Assante y otros proponían un análisis de la información adicional que facilitaban algunos de los más importantes repositorios de datos (*3TU, CSIRO, Dryad, Figshare, Zenodo*). Siguiendo su modelo, podríamos distinguir los siguientes atributos:

- Accesibilidad, esto es, existencia de un DOI o URI para llegar al dataset.
- Datos bibliométricos y/o alométricos. Estadísticas de uso y citación.

Tabla I. Tipos de datos abiertos y formatos de archivo más comunes

Tipo de datos	Formatos de archivo
Audiovisuales	.mp3, .wav, .aif, .mp4, .mov, .avi, .mj2, .flac
Textuales	.txt, .rtf, .bib, .doc, .pdf, .odt, .tex
Imágenes	.jpg, .png, .svg, .tiff
Gráficos 3D	.obj, .stl, .ply
Presentaciones	.ppt, .pptx, .pptm
Geográficos/geoespaciales	.keyhole, .GIS, .gif, .dwg, .shp, .dbf, .mif
Hojas de cálculo	.csv, .xls, .tsv, .tab, .ods
Comprimidos	.zip, .rar, .iso
Estructurados	.mdb, .accdb, .html, .xml

- Cobertura espacio temporal.
- Fecha de creación y subida del dataset.
- Formato de archivo.
- Licencias de uso y reutilización.
- Descripción de contenido.
- Publicación relacionada o vinculada con el conjunto de datos.
- Descripción del proyecto o iniciativa, así como de sus entidades financiadoras.
- Procedencia de los datos.
- Materias (palabras clave, etiquetas, etc.).

Conviene subrayar que el grupo de atributos recogidos bajo el epígrafe "procedencia de los datos", en el que se consignarían las informaciones relativas a los métodos, fuentes e instrumentos empleados para obtener los datos, fue, según su estudio el menos frecuente en el conjunto de repositorios evaluados.

Entre las iniciativas puestas en marcha para desarrollar esquemas de descripción orientados a datos de investigación hay que resaltar la encabezada por el *DataCite Metadata Working Group* en 2016. Este equipo internacional de especialistas ha propuesto un esquema de metadatos – *DataCite Metadata Schema Documentation for the Publication and Citation of Research Data* – para este tipo de registros, haciendo hincapié en la necesidad de una interoperabilidad semántica entre la descripción de los datos que facilitan los distintos recursos. Establece una serie de campos o elementos de descripción obligatorios y otro conjunto cuya inclusión es recomendable o que, en última instancia, considera opcionales. Entre los primeros aparecen, como no podría ser de otro modo, elementos como autor, título, editor, año de publicación, o tipo de recurso. Y dentro de los elementos recomendados, es destacable la incorporación de un campo que, desde nuestra perspectiva, podría ser muy útil en la gestión y consulta de datos abiertos para los investigadores: el denominado genéricamente *Description*, que recogería todo tipo de información adicional que no tuviese cabida en el resto de los elementos y que se subdividiría, entre otros, en los campos Abstract, Métodos, Serie, Tabla de contenidos, Información técnica.

Además de la conveniencia de que se propongan campos propios de descripción para un recurso tan especializado como los datasets, para el objetivo de este trabajo es particularmente relevante que se abra la posibilidad de recuperar los datos abiertos no solo por las características formales

habituales sino por los métodos de investigación que se han practicado para obtenerlos. Este tipo de información puede suponer un criterio de filtro muy útil en las búsquedas, además de ofrecer directamente aclaración sobre un aspecto esencial a la hora de decidir acerca de la reutilización o replicación de los datos en investigaciones subsecuentes. Como ejemplo de recursos que incorporan esta descripción y han adaptado este esquema podemos citar la imprescindible plataforma OpenAIRE (*Open Access Infrastructure for Research in Europe*) (Elbaek y Nielsen, 2013) o el repositorio sobre ecología y medio ambiente KNB (*The Knowledge Network for Biocomplexity*).

Otro de los apartados cruciales en el estudio de los datos abiertos es el que se dedica a analizar su uso e impacto como recurso científico. Como se ha mencionado anteriormente, en esa línea habría que determinar su consideración respecto a su relación con la revista científica en la que se han publicado. López-Borrull (2017) distinguía varias posibilidades: que se publiquen como un objeto de información independiente en un repositorio especializado, que se publiquen junto al propio artículo del que forman parte o que se publiquen en forma de *data paper*. Estos artículos de datos constituyen un formato distinto al tradicional artículo de revista, y, de hecho, han dado lugar a las actuales *data journals* en las que los conjuntos de datos son acompañados por información sobre cómo, cuándo y por qué se han recopilado. La modalidad de publicación y su influencia mutua en el impacto de la revista y los índices de citación y visibilidad de los artículos representa un campo abierto para la investigación bibliométrica y para las políticas de edición científica.

En todo caso, varios estudios han apuntado que compartir los datos abiertos tiene una incidencia positiva en los niveles de citación de los propios artículos vinculados (Pinowar y Vision, 2013; Alexandre-Benavent y otros, 2016; Dorch y otros, 2015). También se ha manifestado la importancia que, en el actual entorno colaborativo, tendrán las métricas alternativas para calibrar el impacto o repercusión de estos datasets. Lo que sí parece claro es que estamos ante un panorama que exige nuevas perspectivas, análisis de los indicadores y afinación de las metodologías de medición, dada la necesidad de conocer cuál está siendo el comportamiento de los investigadores a la hora de reutilizar y citar los datos abiertos que otros comparten, las motivaciones que existen tras esa decisión y la eficacia práctica de las distintas modalidades. Aún estamos en fase experimental en el desarrollo de estas métricas (Pezet Mancebo y otros, 2017), aún no hay consenso total al respecto (Melero y Hernández-San Miguel,

2014; Ball y Duke, 2015) y, a medida que se desarrolla esta práctica, se precisan nuevos análisis también en función del tipo documental, lo que deberá incluir a los *data papers* (Robinson-García y otros, 2016). También ha de abordarse con mayor detalle la relación entre las citas que reciben los conjuntos de datos y el modo en que éstos son compartidos, desde el punto de vista de su preservación y estabilidad como objetos digitales (Peters y otros, 2016).

Por todo lo anterior, la mera disponibilidad de indicadores cuantitativos procedentes de bases de datos y recursos informativos no garantiza su comprensión global, a lo que se suman las particularidades entre disciplinas científicas y la dispersión de las fuentes. En un contexto en que las métricas alternativas también han experimentado un impulso en cuanto a su consideración como termómetro del interés y visibilidad de las publicaciones científicas, el alcance de los conjuntos de datos abiertos debe ser tenido en cuenta como una secuencia continua, que conecte la bibliometría tradicional, los nuevos hábitos de consumo y citación de las publicaciones y las otras dimensiones del uso y comunicación científica. En definitiva, el estudio sobre los datos abiertos de investigación desde todos los ángulos posibles representa un desafío real aún por completar, tanto por parte de los gestores de información científica como por parte de los propios investigadores cuyos datos se insta a compartir.

3. METODOLOGÍA

Contexto y objetivo del estudio

Según todo lo expuesto, el presente trabajo pretende describir una faceta fundamental de los datos abiertos de investigación: la metodología empleada para recopilarlos. Este aspecto enlaza con la necesidad de conocer la naturaleza de estos datos, como elemento determinante en su posterior explotación y que, sin embargo, no suele, en la práctica, incluirse en la descripción que se hace de los mismos (Assante y otros, 2016; Childs y otros, 2014). De manera colateral, se trataba de recoger las métricas de uso ofrecidas, esto es, las relativas a visitas, descargas (Peset y otros, 2017) y citación. Como se ha expresado al inicio, se da la circunstancia de que el repositorio de datos *Figshare* ha incluido recientemente en su colección los datasets procedentes de las revistas de la plataforma SciELO, lo que posibilita el acceso a estos conjuntos de datos mediante la asignación a cada uno de ellos de un *Digital Object Identifier* (DOI). Ofrece, asimismo, acceder a una previsualización en el navegador y atribuirles unas estadísticas de uso que incluyen número de visualizaciones, descargas, citas e información altmétrica.

El repositorio *Figshare* permite compartir contenidos científicos en diversos formatos y tipologías de archivos (pósters, datos, vídeos, presentaciones, códigos informáticos, preprints, etc.). Su mantenimiento desde 2012 está a cargo de la empresa *Digital Science*, es de acceso abierto y los derechos de autor de los archivos subidos están protegidos por las licencias *Creative Commons*. Entre sus fines está el fomento de la cultura de la colaboración científica y la disponibilización y gestión de datos abiertos, lo que facilita desde una interfaz de consulta intuitiva. Cada archivo se acompaña de su descripción de contenido, formato de citación e información sobre su uso (Reed, 2016).

Por su parte, SciELO es una biblioteca digital impulsada por la *Fundação de Amparo à Pesquisa do Estado de São Paulo* (FAPESP) y el Centro Latinoamericano y del Caribe de Información en Ciencias de la Salud (BIREME). Desde hace veinte años ofrece en acceso abierto la consulta a ediciones completas de 1.285 revistas científicas publicadas en América Latina y EL Caribe. El éxito de su implantación y difusión en el ámbito iberoamericano y en sus instituciones científicas se sustenta tanto por la disponibilidad de los artículos como por la inclusión de indicadores bibliométricos.

SciELO ha venido promoviendo la adhesión de las revistas que incluye a las prácticas de ciencia abierta, abogando por la apertura de todos los elementos de la comunicación de las investigaciones; así, desde 2015 exige, para su ingreso y permanencia en la colección, la disponibilización de los datos de investigación utilizados en los artículos (SciELO, 2017). Este proceso debe seguir criterios de registro que garanticen la autoría, el uso y la citación de los datos, así como del artículo correspondiente, para aumentar la replicabilidad de los estudios y la visibilidad de las publicaciones (SciELO 2017, 2014). Esta biblioteca digital ha implementado la identificación y la recomendación de repositorios de datos de investigación por área temática para guiar el depósito de dichos datos. SciELO sigue los mencionados principios FAIR para la gestión de esos repositorios y desde principios de 2018 emplea *Figshare* como repositorio de datos para dar acceso a los datos de unas 200 de sus revistas (solo aquellas con licencia CC-BY, o con otra licencia pero que previamente hayan autorizado la exportación de los datos). Por ahora, la recolección de datos de investigación suplementarios está siendo desarrollada por *Figshare*, de forma automática, a partir de los artículos publicados en SciELO.

Según Hyndman (2018) el repositorio abarca las revistas de la colección SciELO Brasil y en él a cada revista se la ha dotado de una página propia desde la que presentar sus datos. Como parte de este acuerdo de colaboración, todos los datos del por-

tal quedan almacenados de forma segura y reciben un DOI. Posibilita también que todos los tipos de archivo sean visualizados directamente desde el navegador. Además, como se ha mencionado anteriormente, determinadas estadísticas, incluidas el número de visualizaciones, descargas y citas, están disponibles para cada ítem.

La alianza entre estos dos recursos ha habilitado las condiciones para llevar a cabo el presente trabajo, cuyo objetivo fundamental es estudiar la naturaleza y características metodológicas de los datasets procedentes de la colección SciELO disponibles desde *Figshare*.

Recogida de datos y criterios de análisis

La recogida de datos se llevó a cabo en dos momentos (en abril y en octubre de 2018), tomando todos los datasets que provenían de revistas de SciELO. Ello devolvió un total de 361 ficheros de datos, pertenecientes todos ellos a publicaciones brasileñas. De ese conjunto se desestimaron 13 al comprobar que, aunque tenían entradas independientes, estaban duplicados; por tanto, la población de análisis se constituyó finalmente con 348 datasets.

Como punto de partida se tomaron los datos facilitados en la descripción por la propia plataforma, si bien para recabar información sobre algunas características se hizo necesario cotejar individualmente cada uno de los archivos. Con la siguiente relación de variables o criterios de análisis se elaboró una base de datos (Tabla II):

Tabla II. Criterios de análisis aplicado al conjunto de datasets

Título	Fecha de depósito
Número de autores	Revista
Categoría temática	Subcategoría temática
Visitas	Descargas
Citas	Valores alométricos
Metodología	Técnica de obtención
Formato de archivo	Licencia

Los criterios de análisis que no fueron extraídos directamente de la descripción de los archivos fueron los referidos a la metodología empleada (cuantitativa o cualitativa) y la técnica de obtención de datos aplicada en la investigación, siguiendo el esquema de tipologías anteriormente aludido, que distinguía entre datos derivados o compilados, datos experimentales, datos observacionales, datos

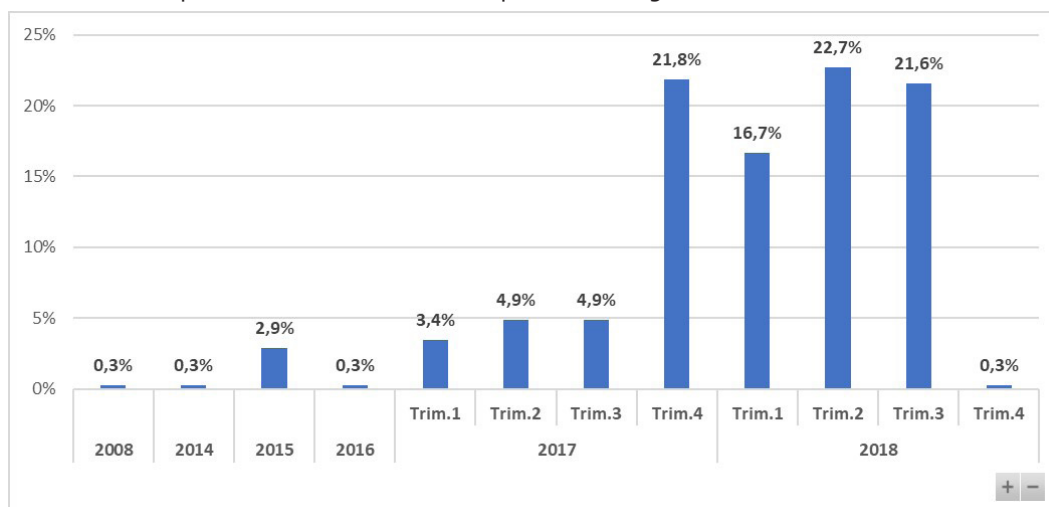
referenciados y datos simulados. Para ese análisis se consultó cada uno de los artículos de los que procedían los datos, confrontando detenidamente sus apartados metodológicos mediante un análisis de contenido. En lo referido al apartado temático, se decidió emplear la propia jerarquización temática asignada por *Figshare*, con el fin de facilitar la comparación de los resultados.

4. RESULTADOS

En cuanto a la fecha de depósito en la plataforma, los datasets analizados se concentraban fundamentalmente entre finales del 2017 y los tres primeros trimestres del 2018, como se puede observar en la Figura 1. Este dato era esperable dado que el acuerdo entre *Figshare* y SciELO está recién estrenado, luego es previsible que las cifras de incorporación de conjuntos de datos sigan aumentando. Con respecto a la fecha de los datasets cabe señalar que esta coincide en todos los casos con la fecha de publicación del artículo del que proceden. Por otro lado, la brevedad del período de tiempo transcurrido puede considerarse un factor que explique la escasez de indicadores de uso y citación posteriores.

Se consideró apropiado desde el punto de vista del análisis de los resultados consignar el número de autores que firmaban los archivos de datos. En este sentido, se contabilizaron un total de 1439 firmas distintas, lo que supone para el conjunto un índice de coautoría de algo más de cuatro autores (4,1) por dataset. La colaboración, que está en la base de la disponibilidad en abierto de los datos de investigación, parece también darse en cuanto a la autoría conjunta de los trabajos más susceptibles a compartir archivos de resultados; así, solo 41 de los 348 datasets analizados estaban firmados por un solo autor. En lo referido a las publicaciones periódicas con mayor número de datasets, la relación de títulos se muestra en la Tabla III, limitada a aquellas revistas que aportaban al menos cuatro archivos de datos.

A continuación, se analizó, para cada una de las revistas identificadas, el nivel de adopción de prácticas encaminadas a compartir sus datos en abierto. Mediante el cotejo de sus políticas de recepción de manuscritos y de las instrucciones a los autores se comprobó que solo tres de ellas aludían a la inclusión de datos de investigación: *Revista Árvore* (que era la que más datasets aportó y la que de un modo más explícito y desarrollado recogía su adhesión a esta práctica), *Trends in Psychiatry and Psychotherapy* y *Arquivos Brasileiros de Oftalmologia*, a cuya colección pertenecían, respectivamente, 7 y 11 de los datasets recopilados.

Figura 1. Fecha de depósito de los datasets en la plataforma *Figshare***Tabla III.** Listado de títulos de revistas y número de datasets aportados

REVISTA	Nº Datasets
Revista Árvore	21
Revista Brasileira de Enfermagem	14
Ciência Rural	13
Arquivos Brasileiros de Oftalmologia	11
Trabalho, Educação e Saúde	9
Revista Brasileira de Educação Médica	8
Ciência e Saúde Coletiva	8
Jornal Brasileiro de Patologia e Medicina Laboratorial	8
Psicologia Escolar e Educacional	8
Revista Brasileira de Ginecologia e Obstetrícia	7
Revista Bioética	7
Rodriguésia	7
Interface-Comunicação, Saúde, Educação	7
Arquivo Brasileiro de Medicina Veterinária e Zootecnia	7
Trends in Psychiatry and Psychotherapy	7
São Paulo Medical Journal	6
Revista da Sociedade Brasileira de Medicina Tropical	6
História, Ciências, Saúde-Marguinhos	6
Memórias do Instituto Oswaldo Cruz	6
Alfa: revista de lingüística	5
Pro-posições	5
Brazilian Oral Research	5
Psico-USF	5
REMHU	4
Brazilian Journal of Otorhinolaryngology	4
Cadernos Saúde Coletiva	4
Revista Brasileira de Estudos de População	4
Transinformação	4

Desde el punto de vista temático, los datos arrojaron una clara prevalencia de las Ciencias de la Salud en cuanto al número de datasets recopilados: el 42,2% pertenecían a distintas áreas médicas. Como se ha señalado, para el tratamiento de los datos se han respetado, traduciendo las denominaciones, las categorías asignadas por *Figshare*. En la figura 2 se expone la relación de las materias más frecuentes y el porcentaje de archivos recopilados. Se observa en este punto que la suma de las áreas médicas y la Biología constituía más de la mitad del total (52,7%), mientras que las áreas que podemos agrupar bajo el epígrafe de Ciencias Sociales y Humanidades representaban casi un 26%. Si desglosamos el grupo mayoritario, las subcategorías más numerosas fueron, por este orden: Medicina General, Salud y Seguridad Laboral, Biotecnología médica, Patología, Neurociencias, Enfermería y Psiquiatría.

El enfoque fundamental de este trabajo era el análisis de la metodología empleada en los datasets tratados. En esa dirección, tras examinar pormenorizadamente las investigaciones de procedencia, se comprobó que en realidad no todo el conjunto recuperado se correspondía efectivamente con archivos de datos de investigación. En algunos casos los datasets solo constituían informaciones, fundamentalmente textuales pero presentadas en forma de tabla, que habían sido depositadas en *Figshare* con formato de hoja de cálculo. Estos ar-

chivos fueron descartados porque no obedecían en ningún caso a la intención de exponer resultados surgidos de aplicar una determinada metodología de investigación sino al propósito de llamar la atención sobre algún aspecto o representar de un modo más legible o sistemático la información descrita. Consiguientemente, fueron descartados 25 datasets (el 7,2% del total) por no contener datos de investigación. El conjunto restante (323) fue clasificado en función de si aportaban datos cuantitativos o cualitativos; los primeros supusieron el 61,3% del total (198 datasets) mientras que los que contenían datos cualitativos sumaron 125 datasets (el 38,7%).

Esa proporción entre el uso de metodologías cuantitativas y cualitativas se presenta también de forma desglosada atendiendo a los resultados en función de la disciplina científica (Figura 3).

El empleo de técnicas de índole cuantitativa era unánime o mayoritario en áreas tecnológicas, Psicología, Biología, Veterinaria y Ciencias Agrícolas, o Ciencias Geológicas y Ambientales, mientras que era mucho más frecuente el uso de métodos cualitativos en Información y Documentación, Ciencias Humanas y Políticas, y Humanidades. Los campos en que el reparto entre ambos enfoques de investigación era más equitativo fueron Ciencias de la Salud, Ingeniería y Ciencias Sociales (aunque en los tres casos, con prevalencia de lo cuantitativo).

Figura 2. Distribución por categorías temáticas de los datasets recuperados

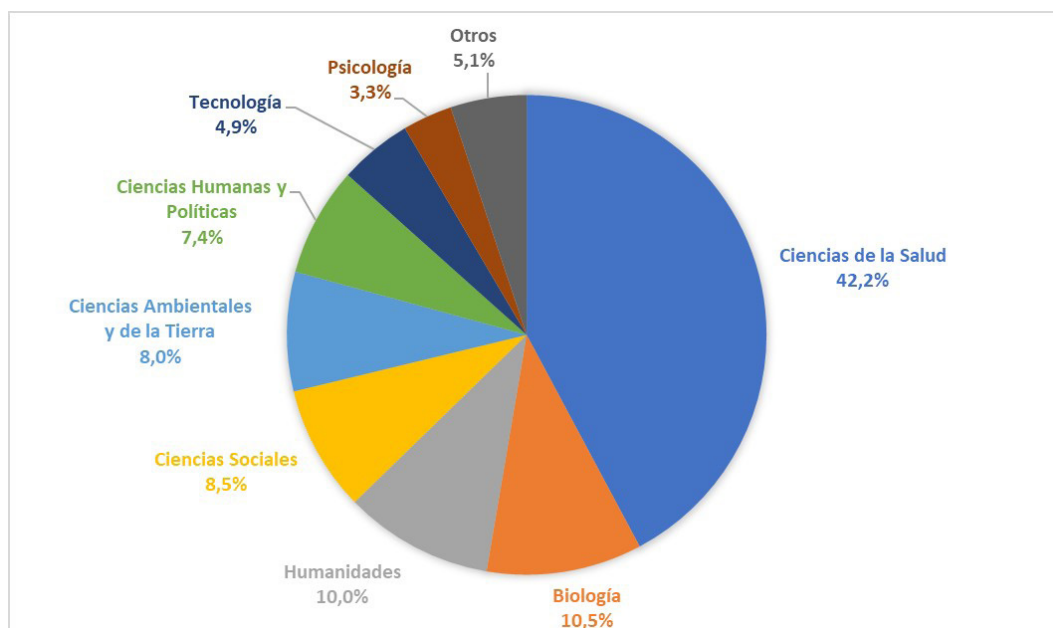
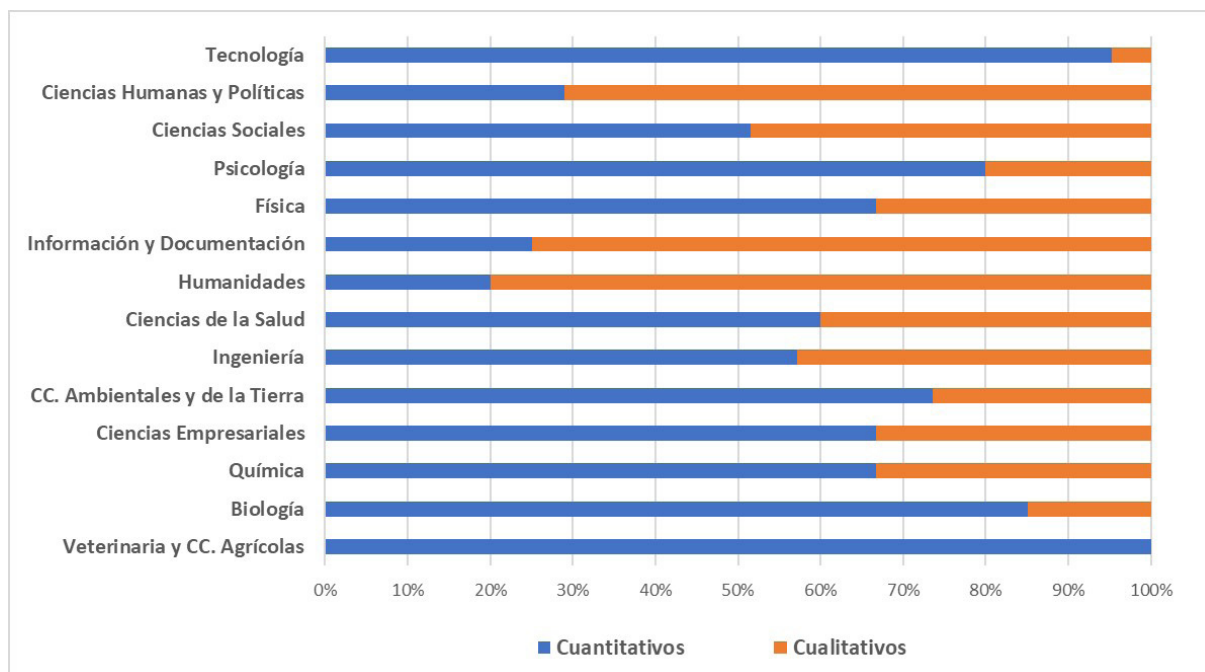


Figura 3. Datos de investigación cuantitativos y cualitativos distribuidos por disciplina científica



Continuando con esta línea de análisis metodológico, se distribuyeron los conjuntos de datos en función de la técnica que había sido empleada para su obtención. Hecha la salvedad sobre el subconjunto que en realidad no contenía datos de investigación originales, el resto se repartía como muestra la figura 4. En ella se plasma una cierta variedad metodológica a la hora de plantear la extracción de datos científicos, pero con una preeminencia de las técnicas observacionales, cuyos resultados se concretan en datos menos replicables en idénticas condiciones.

A continuación, se ofrece el resumen (Tabla IV) de las técnicas de recopilación de datos segregadas en esta ocasión por disciplinas científicas. Se aprecia una notable variedad metodológica para las disciplinas recogidas, pero si atendemos a cada una de ellas cabe destacar, como podría esperarse, que no hay investigaciones que recojan datos experimentales pertenecientes a ninguna de las Ciencias Sociales o Humanidades; estas ramas son mayoritarias entre los datasets que emplearon datos referenciados, esto es, procesaron datos ya publicados en fuentes oficiales. En cuanto a las

Figura 4. Porcentaje de datasets en función de las técnicas de obtención de datos empleadas

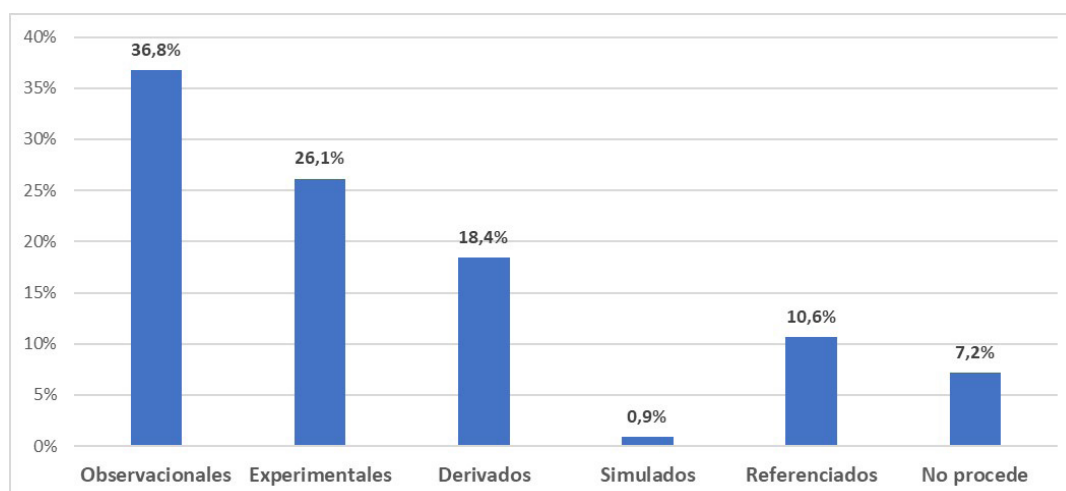


Tabla IV. Datasets según técnica de obtención de datos y disciplina científica

	Nº Datasets	%
Observacionales		
Biología	16	12,5%
Química	2	1,6%
CC. Empresariales	3	2,3%
CC. Ambientales y de la Tierra	11	8,6%
Ingeniería	2	1,6%
Ciencias de la Salud	52	40,6%
Humanidades	7	5,5%
Física	1	0,8%
Psicología	13	10,2%
Ciencias Sociales	5	3,9%
CC. Humanas y Políticas	10	7,8%
Tecnología	6	4,7%
Total	128	100%
Experimentales		
Veterinaria y CC. Agrícolas	1	1,1%
Biología	17	18,7%
Química	1	1,1%
CC. Ambientales y de la Tierra	14	15,4%
Ingeniería	1	1,1%
Ciencias de la Salud	42	46,2%
Psicología	1	1,1%
Tecnología	14	15,4%
Total	91	100%
Derivados		
Biología	1	1,6%
Ingeniería	4	6,3%
Ciencias de la Salud	25	39,1%
Humanidades	13	20,3%
Información y Documentación	3	4,7%
Psicología	1	1,6%
Ciencias Sociales	6	9,4%
CC. Humanas y Políticas	10	15,6%
Tecnología	1	1,6%
Total	64	100%
Simulados		
CC. Ambientales y de la Tierra	2	66,7%
Física	1	33,3%
Total	3	100%

	Nº Datasets	%
Referenciados		
Biología	2	5,4%
CC. Ambientales y de la Tierra	1	2,7%
Ciencias de la Salud	10	27,0%
Humanidades	6	16,2%
Información y Documentación	1	2,7%
Ciencias Sociales	7	18,9%
CC. Humanas y Políticas	10	27,0%
Total	37	100%
No procede		
	25	

que hicieron uso de la observación, la técnica más frecuente si tomamos todos los casos en conjunto, se registró gran variedad de campos científicos, abarcando todo el espectro temático. Los datos de investigación obtenidos por derivación de compilaciones de datos anteriores fueron los más comunes en Ciencias Humanas y Políticas. Por último, como era previsible por la gran representación en la muestra analizada de datasets pertenecientes a Ciencias de la Salud, dicha materia estuvo presente en modo preponderante en casi todas las técnicas de recogida de datos.

Además de la metodología aplicada en los datasets, otro de los aspectos que se quería explorar eran los indicadores de uso y citación. En cuanto a los de uso, concretados en el número de visitas y descargas contabilizados para cada archivo, las cifras fueron poco significativas. El conjunto de los datasets solo se había visitado 2.187 veces y descargado en 406 ocasiones. Estas cifras son de por sí bajas (con medias de 6,3 visitas y apenas 1,1 descargas por documento); pero estos índices son en realidad todavía más reducidos, dado que se comprobó la existencia de cuatro datasets que por sí solos acaparaban más de la mitad de las visitas (1.149) y más de la mitad de las descargas (223). Curiosamente las materias de esos documentos tan usados pertenecían a materias que no estaban entre las más frecuentes (Ciencias Humanas y Políticas, Información y Documentación, y Humanidades). En lo que se refiere a citas recogidas, todo el conjunto solo reunía una cita bibliográfica consignada. En definitiva, los indicadores de uso y citación pueden considerarse poco relevantes en esta colección a la hora de establecer consideraciones o extrapolaciones futuras, pero puede constituir un punto de partida sobre el que descifrar las razones

que explican esa realidad. Ello redunda en lo apuntado por estudios anteriores en el sentido de que aún es preciso resolver cuestiones metodológicas antes de hacer significativas estas métricas de uso y citación para conjuntos de datos (Robinson-García y otros, 2016; Peters y otros, 2016).

Por último, los dos últimos criterios que se analizaron del conjunto, esto es, formato de archivo y licencia de uso, arrojaron resultados idénticos para el total de datasets: todos estaban en formato *.xls* y a todos les correspondía la licencia *CC by 4.0*. Si bien la homogeneidad en este último aspecto tiene su justificación en la misma procedencia de todos los documentos (la propia colección SciELO) con políticas sobre derechos de uso semejantes, se encontró menos oportuno que todos los archivos tuviesen el formato de hoja de cálculo dada la variedad de tipos de datos identificados.

5. DISCUSIÓN Y CONCLUSIONES

A la vista de los resultados expuestos, una de las consideraciones más importantes que pueden plantearse es la conveniencia de ampliar la información que acompaña a los propios datos de investigación. Como han señalado otros autores (Thanos, 2015), la descripción de los datasets es un factor determinante para fomentar el uso posterior de los datos de investigación. Y dada la heterogeneidad potencial de los datos de investigación, su especificidad, su proliferación y la diversidad de sistemas y herramientas de búsqueda, parece evidente la necesidad de facilitar representaciones lo más completas posible de los mismos. Sin esa descripción detallada y normalizada será más complicado el descubrimiento de estos recursos y consecuentemente menguará su reutilización futura.

También supone un aspecto de incidencia directa la forma de hacer accesibles en abierto los datos de investigación. Entre las distintas opciones posibles, el conjunto analizado en este estudio comparte una misma característica: está depositado en un repositorio, pero cada registro está vinculado al artículo del que procede. A partir del convenio de colaboración entre *Figshare* y SciELO, estamos pues ante un modelo peculiar, a medio camino entre el depósito directo en un repositorio y la publicación desde la propia publicación periódica.

Desde esa perspectiva, es oportuno tener en cuenta el estudio de las políticas adoptadas por las propias revistas para facilitar los datasets de los trabajos que publican. En el trabajo de Sturges y otros (2014) ya se ponía de manifiesto que las publicaciones científicas no eran del todo precisas a la hora de plantear a los autores su estrategia de publicación de datos en abierto y se alentaba a concretar esa circunstancia. Actualmente las revistas están desarrollando protocolos más rigurosos para compartir los datos de investigación, pero, en el caso que nos ocupa nos encontramos ante una colección de títulos muy variada, donde el modo empleado ha sido transformar en un formato tabulado, de hoja de cálculo, los datos aportados por los artículos de procedencia, independientemente de la metodología empleada para obtenerlos o de su naturaleza. De hecho, del conjunto de revistas incluidas, solo tres habían enunciado de forma expresa alguna mención a su política de datos abiertos. Si bien la incorporación a *Figshare* constituye un esfuerzo novedoso y significativo, cabe esperar que, en próximas iniciativas, las propias revistas establezcan un estándar de procedimiento que delimite un abanico más rico de formatos posibles y especifique los metadatos que acompañan a esos datasets, así como unos requisitos de validez mínimos a cumplir por los mismos. Ello evitaría, por ejemplo, uno de los hallazgos de este estudio, en el que casi un 7,2% de los conjuntos de datos estudiados no eran, en sentido estricto, datos de investigación.

En lo referido al objetivo de análisis principal de este trabajo, la metodología empleada en los conjuntos de datos observados, una de las condiciones que se quiere poner de relieve es la necesidad de definir de forma homogénea en los propios trabajos de investigación los métodos y técnicas empleados. Aunque es un aspecto que excede el ámbito de la ciencia abierta, tras el análisis individualizado de los trabajos incluidos en este estudio, cabe hacer una reflexión sobre la pertinencia de mejorar la forma en que se exponen las decisiones metodológicas y las estrategias de recogida y análisis de los datos en los

artículos científicos. Partiendo de un escenario donde la variedad y la heterogeneidad es la norma y la interdisciplinariedad va en aumento, la delimitación metodológica es fundamental para calibrar la utilidad de un archivo de datos. En determinadas áreas, como en Información y Documentación (Guallar y otros, 2017) son habituales los estudios que hacen hincapié en el abordaje metodológico de sus propios artículos, pero este enfoque se hace aún más apremiante al tratar conjuntos de datos abiertos, donde la metodología para su obtención es determinante para su validez y significación. Prueba de este interés por normalizar la exposición metodológica de los trabajos de investigación y de los datos resultantes es, por ejemplo, la reciente aparición de publicaciones amparadas por los grupos de trabajo de la *American Psychological Association* (Appelbaum y otros, 2018; Levitt y otros, 2018).

En trabajos anteriores, como los de Akers y Doty (2013) o Van Tuyl y Michalek (2015), se ha abordado la cuestión de la predisposición de los autores a depositar sus datos de investigación en acceso abierto y el conocimiento del proceso y su regulación. Esta aproximación basada en encuestas y/o entrevistas ha puesto de manifiesto diferencias en función de las disciplinas, planteando, por ejemplo, que los investigadores de áreas médicas son los más reacios a compartir sus datos de investigación. En el caso del presente estudio, sin embargo, han sido precisamente los datasets de materias relativas a Ciencias de la Salud los más numerosos. En cualquier caso, hay que tener en cuenta que esta distribución ha venido más determinada por las características temáticas de las revistas incluidas en la colección en SciELO que por la voluntad explícita de los investigadores o autores implicados.

Con los resultados obtenidos se ha revelado un mayor uso de técnicas cuantitativas frente al uso de técnicas cualitativas de investigación si tomamos todos los datasets de forma conjunta (61,3% y 38,7%, respectivamente), pero si observamos cada disciplina, se ven diferencias en cuanto a la prevalencia de cada metodología. Cabe apuntar, como línea de investigación futura, la comparación y correspondencia de estos datos con las metodologías de investigación habituales de cada área temática. También es adecuado reseñar que para datos cualitativos pueden darse casos en los que la disponibilidad en abierto de los datos es más susceptible de entrar en conflicto con algunos aspectos tanto éticos como de replicabilidad de los mismos (Childs y otros, 2014). Desde el punto de vista de la obtención de los datos, se apreció variedad en el empleo de las técnicas, pero los datos resultantes de la observación fueron los más numerosos, se-

guidos de los obtenidos mediante experimentación y de los derivados de datos anteriores. De nuevo en este apartado se apreciaron diferencias en función del campo temático, corroborando la tendencia clásica a identificar los métodos experimentales con las ciencias naturales. De cualquier modo, es notorio en este punto la pluralidad de métodos empleados tanto en Ciencias de la Salud como en las Ciencias Sociales.

En cuanto al uso, parece claro que para esta colección el cómputo de indicadores es aún prematuro. Al hilo de lo que algunos autores ya apuntaban, respecto a que estamos en una fase inicial de desarrollo de métodos para medir el acceso y consumo de los datos (Peset Mancebo y otros, 2017), en el presente estudio no se cuenta con indicios suficientes para correlacionar el número de visitas y descargas con el número de citas recibidas, en parte debido a la ventana temporal de los datasets incluidos. También se han detectado posibles inconsistencias en la medición de los indicadores de visitas y descargas para algunos datasets que aparecen sobrerrepresentados, lo que enlaza con la necesidad de establecer filtros de calidad, fiabilidad y precisión en el cálculo de estas alométricas.

Tras todo lo anterior, se pueden plantear varias cuestiones que se estima incidirán en el análisis futuro de los conjuntos de datos abiertos:

- la delimitación conceptual del impacto de los datos de investigación, tomando las distintas dimensiones y métricas tradicionales y alternativas (Ball y Duke, 2015) y su relación con el resto de productos de la publicación científica (Onyancha, 2016; Aleixandre-Benavent y otros, 2016).
- el seguimiento y comparación de las nuevas herramientas y recursos que recogen datasets y su procedencia.

- la normalización de la forma de citación de los datos de investigación como producto científico.
- la validación o verificación desde los propios editores o repositorios de los conjuntos de datos a los que dan acceso, no solo para asegurar que son efectivamente datos de investigación originales, sino como parte de un ciclo de revisión adicional que garantice la idoneidad de reutilizar esos datasets en investigaciones siguientes.

La reutilización efectiva de los datos científicos en abierto pasa necesariamente no solo por la adhesión a esta práctica por parte de los autores, sino también por la consideración de estos conjuntos de datos por parte de las distintas agencias y entidades de evaluación de la investigación. Ello, junto a un impulso en la mejora de la descripción analítica que acompaña a estos recursos, puede facilitar su extensión como hábito de comunicación real también en áreas donde siguen siendo menos frecuentes.

Como se ha comprobado, aún no se ha generalizado como norma la inclusión en los metadatos de los datasets de la información referida a técnicas de investigación empleadas y procesos de recogida y tratamiento de los datos. Consecuentemente, y a pesar de la dificultad intrínseca de estos estudios enfocados en el análisis de la metodología, se recomienda profundizar en esta línea de investigación, atendiendo a las diferencias entre disciplinas y a las particularidades de ámbitos como el iberoamericano, tradicionalmente menos representados en las principales bases de datos científicas. En un contexto donde la investigación es cada vez más híbrida y multidisciplinar, la mejora en la caracterización de los datos abiertos ayudará, por un lado, al reconocimiento de su fiabilidad y rigor científico y, por otro, facilitará recuperaciones más eficientes en un entorno de recursos interoperables.

6. REFERENCIAS

Akers, K. G.; Doty, J. (2013). Disciplinary differences in faculty research data management practices and perspectives. *International Journal of Digital Curation*, 8 (2), 5-26. <http://dx.doi.org/10.2218/ijdc.v8i2.263>

Aleixandre-Benavent, R.; Moreno-Solano, L.; Ferrer Sapeña, A.; Sánchez Pérez, E. A. (2016). Correlation between impact factor and public availability of published research data in Information Science and Library Science journals. *Scientometrics*, 107 (1), 1-13. <http://doi.org/10.1007/s11192-016-1868-7>

Appelbaum, M.; Cooper, H.; Kline, R. B.; Mayo-Wilson, E.; Nezu, A. M.; Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, 73 (1), 3-25. <http://dx.doi.org/10.1037/amp0000191>

Assante, M.; Candela, L.; Castelli, D.; Tani, A. (2016). Are Scientific Data Repositories Coping with Research Data Publishing? *Data Science Journal*, 15 (6). <http://doi.org/10.5334/dsj-2016-006>

- Ayris, P. (2016). Supporting the uptake of Research Data Management (RDM): Introducing the LEARN Project. *DLib Magazine*, 22 (7/8). <http://www.dlib.org/dlib/july16/07inbrief.html>
- Ball, A.; Duke, M. (2015). *How to Track the Impact of Research Data with Metrics. DCC How-to Guides*. Edinburgh: Digital Curation Centre. Disponible en: <http://www.dcc.ac.uk/resources/how-guides/track-data-impact-metrics> [Fecha de consulta: 22/10/2018].
- Beagrie N.; Houghton J.W. (2014) *The Value and Impact of Data Sharing and Curation: A synthesis of three recent studies of UK research data centres*. Joint Information Systems Committee (Jisc). Disponible en: http://repository.jisc.ac.uk/5568/1/iDF308_-_Digital_Infrastructure_Directions_Report%2C_Jan14_v1-04.pdf [Fecha de consulta: 15/09/2018].
- Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the Association for Information Science Technology*, 63 (6), 1059-1078. <http://doi.org/10.1002/asi.22634>
- Candela, L.; Castelli, D.; Manghi, P.; Tani, A. (2015). Data Journals: A Survey. *Journal of the Association for Information Science Technology*, 66 (9), 1747-1762. <https://doi.org/10.1002/asi.23358>
- Childs, S.; McLeod, J.; Lomas, E.; Cook, G. (2014). Opening research data: issues and opportunities", *Records Management Journal*, 24 (2), 142-162. <https://doi.org/10.1108/RMJ-01-2014-0005>
- DataCite Metadata Working Group (2016). *DataCite Metadata Schema Documentation for the Publication and Citation of Research Data. Version 4.0*. DataCite e.V. <http://doi.org/10.5438/0012>.
- Dorch, S. B.; Drachen, T.; Ellegaard, O. (2015). The data sharing advantage in astrophysics. *Proceedings of the International Astronomical Union*, 11, 172-175. <https://doi.org/10.1017/S1743921316002696>
- Elbaek, M.; Nielsen, L. H. (2013). *OpenAIRE Guidelines for Data Archive Managers v1.0*. Zenodo. <http://doi.org/10.5281/zenodo.6918>
- Estevão, J. S. B.; Arns, E. M.; Strauhs, F. R. (2017). A ciência aberta na gestão de dados de pesquisa: uma abordagem para abrir a caixa preta da produção científica. En: *XVII Congresso Latino-Ibero-americano de Gestão Tecnológica*, Ciudad de Mexico: ALTEC. http://www.uam.mx/altec2017/pdfs/ALTEC_2017_paper_146.pdf [Fecha de consulta: 22/10/2018].
- European Commission. (2016). *Guidelines to the rules on Open Access to Scientific Publications and Research Data in Horizon 2020*. Disponible en: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf [Fecha de consulta: 10/10/2018].
- European Commission (2016b) *Open Innovation, Open Science, Open to the World - A Vision for Europe*. Disponible en: <https://ec.europa.eu/digital-single-market/en/news/open-innovation-open-science-open-world-vision-europe> [Fecha de consulta: 10/10/2018].
- Fecher, B.; Friesike, S. (2014). *Open Science: One Term, Five Schools of Thought*. En: *Opening Science*. Amsterdam: Springer. http://doi.org/10.1007/978-3-319-00026-8_2
- FECYT- Grupo de Trabajo de "Depósito y Gestión de datos en Acceso Abierto" del proyecto RECOLECTA (2012). *La conservación y reutilización de los datos científicos en España. Informe del grupo de trabajo de buenas prácticas*. Madrid: Fundación Española para la Ciencia y la Tecnología, FECYT. Disponible en: <http://digital.csic.es/handle/10261/65317> [Fecha de consulta: 15/06/2018].
- García-García, A.; López-Borrull, A.; Peset, F. (2015). Data journals: eclosión de nuevas revistas especializadas en datos. *El profesional de la información*, 24 (6), 845-854. <https://doi.org/10.3145/epi.2015.nov.17>
- Guallar, J.; Ferran-Ferrer, N.; Abadal, E.; Server, A. (2017). Revistas científicas españolas de información y documentación: análisis temático y metodológico. *El profesional de la información*, 26 (5), 947-960. <https://doi.org/10.3145/epi.2017.sep.16>
- Hyndman, A. (2018). Figshare and SciELO Partner to Bring Supplementary Data to the Forefront. *Figshare Blog*, jan 16, 2018. Disponible en: https://figshare.com/blog/Figshare_and_SciELO_Partner_to_Bring_Supplementary_Data_to_the_Forefront/363 [Fecha de consulta: 10/10/2018].
- Kindling, M.; Pampel, H.; van de Sandt, S.; Rücknagel, J.; Vierkant, P.; Kloska, G.; Witt, M.; Schirmbacher, P.; Bertelmann, R.; Scholze, F. (2017). The Landscape of Research Data Repositories in 2015: A re3data Analysis. *D-Lib Magazine*, 23 (3/4). <http://doi.org/10.1045/march2017-kindling>
- Krier, L.; Strasser, C.A. (2014). *Data management for libraries: a LITA guide*. Chicago, American Library Association.
- Levitt, H. M.; Bamberg, M.; Creswell, J. W.; Frost, D. M.; Josselson, R.; Suárez-Orozco, C. (2018). Journal article reporting standards for qualitative primary, qualitative meta-analytic, and mixed methods research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, 73 (1), 26-46. <https://doi.org/10.1037/amp0000151>
- López-Borrull, A. (2017). Cambios y tendencias en la publicación de revistas científicas. En: Abadal, E. (ed.), *Revistas científicas: situación actual y retos de futuro*, pp. 221-237. Barcelona: Universitat de Barcelona.
- Marcial L.H.; Hemminger B.M. (2010). Scientific Data Repositories on the web. An initial survey. *Journal of the American Society for Information Science and Technology*, 61 (10), 2029-2048. <https://doi.org/10.1002/asi.21339>
- Melero, R.; Hernández-San-Miguel, J. (2014). Acceso abierto a los datos de investigación, una vía hacia la colaboración científica. *Revista Española de Documentación Científica*, 37 (4): e066. doi: <http://dx.doi.org/10.3989/redc.2014.4.1154>

- OCDE (2015). Making Open Science a Reality, *OECD Science, Technology and Industry Policy Papers*, 25, OECD Publishing, Paris. <https://doi.org/10.1787/5jrs2f963zs1-en>
- Onyancha, O.B. (2016). Open Research Data in Sub-Saharan Africa: A Bibliometric Study Using the Data Citation Index. *Publishing Research Quarterly*, 32 (3), 227-246. <https://doi.org/10.1007/s12109-016-9463-6>
- Peset, F.; Fernández-López, A. (2014). Carencias informativas de los datos abiertos en España. *Anuario ThinkEPI*, 8, pp. 318-321.
- Peset Mancebo, F.; Aleixandre Benavent, R.; Blasco Gil, Y; Ferrer Sapena, A. (2017). Datos abiertos de investigación. Camino recorrido y cuestiones pendientes. *Anales de Documentación*, 20 (1). <https://doi.org/10.6018/analesdoc.20.1.272101>
- Peters, I.; Kraker, P.; Lex, E.; Gumperberger, C.; Gorraiz, J. (2016). Research data explored: an extended analysis of citations and altmetrics. *Scientometrics* 107 (2), 723-744. <https://doi.org/10.1007/s11192-016-1887-4>
- Piowar, H. A.; Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, 1, e175 <https://doi.org/10.7717/peerj.175>.
- Ramjoué, C. (2015). Towards Open Science: The visión of the European Commission. *Information Services & Use*, 35 (3), 167-170. <https://doi.org/10.3233/ISU-150777>
- Reed, R. B. (2016). Figshare for Institutions. *Journal of the Medical Library Association*, 104 (4), 376-378. <https://doi.org/10.3163/1536-5050.104.4.031>
- Ríos-Hilario, A. B.; Travieso-Rodríguez, C. (2013). Estudio de la Revista Española de Documentación Científica: análisis comparativo de los periodos 1989-1999 y 2000-2010. *Revista Española de Documentación Científica*, 36 (3), e016. <https://doi.org/10.3989/redc.2013.3.990>
- Robinson; García, N.; Jiménez; Contreras, E.; Torres; Salinas, D. (2016). Analyzing data citation practices using the data citation index. *Journal of the Association for Information Science and Technology*, 67, 2964-2975. <https://doi.org/10.1002/asi.23529>
- SciELO (2014). *Crítérios, política e procedimentos para a admissão e a permanência de periódicos científicos na Coleção SciELO Brasil*. São Paulo: SciELO. Disponible en: http://www.scielo.br/avaliacao/20141003NovosCriterios_SciELO_Brasil.pdf [Fecha de consulta: 10/10/2018].
- SciELO (2017). *Crítérios, política e procedimentos para a admissão e a permanência de periódicos científicos na Coleção SciELO Brasil*. São Paulo: SciELO. Disponible en: http://www.scielo.br/avaliacao/20171206NovosCriterios_SciELO_Brasil.pdf [Fecha de consulta: 15/10/2018].
- Stony Brook University Libraries. (2018). *Data Management Topics*. Disponible en: <http://guides.library.stonybrook.edu/research-data-services>. [Fecha de consulta: 11/09/2018].
- Sturges, P.; Bamkin, M.; Anders, J.; Hussain, A. (2014). Access to Research Data: Addressing the Problem through Journal Data Sharing Policies. *Proceedings of the IATUL Conferences*. <http://docs.lib.purdue.edu/iatul/2014/openaccess/3>
- Thanos, C. (2015). Scientific Data Reusability: Concepts, Impediments and Enabling Technologies. V *International Conference on Digital Presentation and Preservation of Cultural and Scientific Heritage-DIPP2015*, pp. 19-30.
- Thomson Reuters. (2012). *Repository evaluation, selection, and coverage policies for the Data Citation Index within Thomson Reuters Web of Knowledge*. Disponible en: http://wokinfo.com/products_tools/multidisciplinary/dci/selection_essay/ [Fecha de consulta: 11/06/2018].
- Torres-Salinas, D.; Robinson-García, N.; Cabezas-Clavijo, Á. (2012). Compartir los datos de investigación: introducción al data sharing. *El profesional de la información*, 21 (2), 173-184. <https://doi.org/10.3145/epi.2012.mar.08>
- UN Data Revolution. (2014). *A world that counts: Mobilizing the data revolution for sustainable development*. United Nations: Independent Expert Advisory Group on a Data Revolution for Sustainable Development. Disponible en: <http://www.undatarevolution.org/report/> [Fecha de consulta: 05/10/2018].
- Van Tuyl, S.; Michalek, G. (2015). Assessing Research Data Management Practices of Faculty at Carnegie Mellon University. *Journal of Librarianship and Scholarly Communication*, 3 (3), eP1258. <https://doi.org/10.7710/2162-3309.1258>
- Vicente-Sáez, R; Martínez-Fuentes, C. (2018). Open Science now: A systematic literature review for an integrated definition. *Journal of Business Research*, 88, 428-436. <http://doi.org/10.1016/j.jbusres.2017.12.043>
- Victoria University, Melbourne. (2015). *Research Data Management: data deposit requirements of selected science journals*. Disponible en: <http://libraryguides.vu.edu.au/research-data-management/publications-requiringdatadeposit> [Fecha de consulta: 15/06/2018].