

# Relevamiento de repositorios y portales de datos abiertos de investigación en la Argentina



Camila Indart

Universidad de Buenos Aires. Facultad de Filosofía y Letras. Departamento de Bibliotecología y Ciencia de la Información. Buenos Aires, Argentina. Universidad Nacional de Quilmes. Departamento de Ciencias Sociales. CONICET. Bernal, Buenos Aires, Argentina | camilaindart@gmail.com / <https://orcid.org/0000-0003-2840-3840>

Ana Belén Ríos Hilario

Universidad de Salamanca. Departamento de Biblioteconomía y Documentación. España | [anarihi@usal.es](mailto:anarihi@usal.es) / <https://orcid.org/0000-0002-7052-5860>

Blanca Rodríguez Bravo

Universidad de León. Área de Biblioteconomía y Documentación. España | [blanca.rodriguez@unileon.es](mailto:blanca.rodriguez@unileon.es) / <https://orcid.org/0000-0002-9476-7602>

## Resumen

El artículo presenta el relevamiento realizado con el fin de localizar los portales y repositorios que contienen datos primarios de investigación en la Argentina. Para llevarlo a cabo, se consultaron directorios y registros de repositorios, tanto de datos como de publicaciones, se revisaron los repositorios incluidos en el Sistema Nacional de Repositorios Digitales y aquellos cosechados por el Sistema de Información Universitaria del Consejo Interuniversitario Nacional en su servicio SIU-BDU 2. Además, se realizaron búsquedas en el motor de búsqueda Google y se hicieron relevamientos de fuentes documentales sobre la temática. Se hallaron 25 repositorios y portales que contienen datos y se analizaron algunas características generales de estos.

### Palabras clave

Acceso Abierto  
Datos de investigación  
Repositorios de datos  
Argentina

## Survey of open research data portals and repositories in Argentina

## Abstract

This article presents the survey to locate the portals and repositories that contain primary research data in Argentina. For this purpose, directories and repositories records, both for data and publications, the repositories included in the Sistema Nacional de Repositorios Digitales, and those harvested by the Sistema de Información Universitaria del Consejo Interuniversitario Nacional in its SIU-BDU 2 service were consulted. Besides, Google search engine and documentary sources on the subject were used. We found 25 repositories containing data and we analyzed some general their characteristics.

### Keywords

Open Access  
Research Data  
Data Research Repositories  
Argentina

Artículo recibido: 02-08-2021. Aceptado: 22-11-2021

## 1. Introducción

El movimiento de Acceso Abierto propone un modelo de comunicación científica en el que la producción de los investigadores, cuyos mayores sostenedores son los Estados, esté disponible de forma libre y gratuita para la comunidad internacional a través de plataformas digitales accesibles mediante la web. Los datos primarios de investigación también son objeto de este movimiento, pero presentan un grado de complejidad mayor que las publicaciones para su tratamiento y apertura. Es por esto que representan un desafío para la generación de infraestructuras que soporten esta complejidad y sean de utilidad para la comunidad científica.

En la Argentina, en 2013, se aprobó la Ley 26.899 que obliga a los organismos de ciencia y tecnología que reciben financiamiento del Estado a crear repositorios digitales propios o compartidos con otras instituciones que alojen los resultados de la investigación. Abarca tanto a las publicaciones en que se comuniquen los resultados –tesis, artículos, libros, informes– como a los datos primarios y se establecen plazos de adecuación a la misma.

Con el objetivo de conocer en qué estado se encuentra la apertura de los datos primarios de investigación a la luz de la legislación argentina, se realizó un relevamiento para localizar los repositorios y portales de datos que se encuentran actualmente en funcionamiento. Este trabajo surge a partir del Trabajo Final del Máster en Sistemas de Información Digital de la Universidad de Salamanca defendido en julio de 2020 y representa una actualización del relevamiento luego de un año. Se estructura en 4 partes: descripción de la metodología seguida para realizar el relevamiento; caracterización de los datos abiertos de investigación y breve descripción de las políticas de Acceso Abierto en la Argentina; relevamiento y descripción del conjunto de repositorios hallados; y conclusiones.

## 2. Metodología

La investigación tuvo un carácter descriptivo-exploratorio ya que se propuso trazar un panorama del fenómeno en la Argentina a través del relevamiento de portales y repositorios que alojen datos de investigación. Existen directorios o registros de este tipo de herramientas que permiten conocer a nivel global la existencia de repositorios de diversos tipos, entre los que se cuentan los de datos. El problema identificado cuando se comenzó a realizar este trabajo es la existencia de repositorios de datos de instituciones argentinas que no están inscritos en estos registros. Por esta razón, fue necesario realizar un relevamiento sistemático que intente dar cuenta del universo de portales y repositorios de datos de investigación en la Argentina.

Los pasos para la realización de este relevamiento fueron:

1. Disponible en: <https://repositoriosdigitales.mincyt.gob.ar/vufind/Content/repos>

2. Disponible en: <http://co-sechador.siu.edu.ar/bdu3/>

1. Revisión de registros y directorios de repositorios, tanto de datos como de publicaciones.
2. Revisión de cada uno de los repositorios registrados en el Sistema Nacional de Repositorios Digitales (SNRD)<sup>1</sup> del Ministerio de Ciencia, Tecnología e Innovación (MINCYT).
3. Revisión de cada uno de los 55 repositorios que integran el Sistema de Información Universitaria del Consejo Interuniversitario Nacional (SIU-BDU 2)<sup>2</sup>.
4. Búsquedas en el buscador Google por las palabras claves:
  - a. portal datos abiertos investigación argentina
  - b. repositorio datos abiertos investigación argentina

- c. repositorio datos abiertos ciencia y tecnología argentina
- d. portal datos abiertos ciencia y tecnología argentina

5. Revisión de la literatura producida sobre el tema en la Argentina en busca de iniciativas.

### 3. Acceso Abierto a los datos de investigación

La Organisation for Economic Co-operation and Development (OECD) (2007) define a los datos de investigación como:

Factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings. A research data set constitutes a systematic, partial representation of the subject being investigated (p. 13).

Torres Salinas, Robinson García y Cabezas Clavijo indican que hay un gran consenso en definir a los datos de investigación como “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, Robinson García y Cabezas Clavijo, 2012: 175). La naturaleza de estos registros es variada y depende del campo disciplinar en el que se inscriben: pueden consistir en números, en texto, ser audiovisuales, digitales o físicos (Aleixandre-Benavent, Ferrer Sapena y Peset, 2021).

La clasificación de los datos de investigación de la National Science Board (2005) de Estados Unidos los agrupa en:

- » Datos observacionales: son registros históricos, obtenidos en un lugar y tiempo determinado. Esta característica hace que no puedan ser reproducidos en caso de pérdida. Ejemplos: encuestas de opinión, datos climatológicos.
- » Datos experimentales: surgidos en la planificación y realización de experimentos. Pueden reproducirse, pero el costo de repetir el experimento es alto. Se generan con instrumentos especializados.
- » Datos computacionales: son resultado de ejecutar un modelo computacional. La reproducción requiere documentación del hardware, software y datos de entrada.

El avance acelerado en tecnologías de la información y la comunicación ha dado lugar a proyectos científicos a gran escala, en colaboración internacional y con un uso intensivo de datos. La posibilidad de almacenar grandes volúmenes de datos, la facilidad de su transmisión y la capacidad para procesarlos permitió llevar adelante proyectos tales como el del genoma humano, en el que la información se comparte de forma abierta entre investigadores de distintos países (OECD, 2007). Sin embargo, no solo las *bigsciences* -astronomía, física, genética- se vieron beneficiadas por estos cambios tecnológicos, sino que también los proyectos científicos más pequeños en escala que, a su vez, son los más numerosos (la “larga cola”), se encontraron con la posibilidad de reunir sus datos y ampliar sus investigaciones a partir de ello (Borman, 2012). De esta manera, la disponibilidad de datos en soporte electrónico permite su reutilización en contextos ajenos al de surgimiento, tanto geográficos como disciplinares, así también en el ámbito de la industria y los servicios. De ahí que los grandes volúmenes de datos producidos y almacenados en soporte electrónico se han convertido en un recurso económico de relevancia en las sociedades.

Sin embargo, Borgman (2015) sostiene que los datos científicos no tienen valor ni significado de forma aislada, sino que forman parte de una infraestructura de conocimiento formada por personas, prácticas, tecnologías, instituciones, objetos materiales y relaciones. Señala que el hecho de reconocer un fenómeno como un dato constituye en sí mismo un acto científico. Los datos se construyen y se utilizan en comunidades de práctica que los definen y los valoran como tales en base al conocimiento construido dentro de la misma comunidad. La recolección, el análisis, la interpretación y la gestión de los datos de investigación forman parte de los conocimientos que circulan en un campo científico. Es por esto que, en su trabajo, Borgman enfatiza que poseer un gran volumen de datos y la tecnología para explotarlos no sortea la necesidad de que estos sean representados de manera contextualizada, es decir, que incluyan información sobre en qué contexto y cómo fueron generados para que puedan ser reutilizados.

Las declaraciones que sentaron las bases del movimiento de Acceso Abierto, la Declaración de Budapest, la de Bethesda y la de Berlín, ya incluían a los datos de investigación, si bien en un primer momento se aunaron los esfuerzos en torno a las publicaciones científicas (Melero y Hernández San Miguel, 2014). De la misma manera, organismos multilaterales como la OECD, la Unión Europea y la Comisión Económica para América Latina y el Caribe (CEPAL) comenzaron a generar recomendaciones en esta línea. La OECD organizó en 2004 una reunión de ministros de Ciencia y Tecnología de sus países miembros y como resultado de ella se publicaron una serie de recomendaciones para la publicación de los datos de investigaciones financiadas con fondos públicos (OECD, 2007). Más recientemente elaboró recomendaciones para alcanzar la ciencia abierta donde se enfoca mayormente en el acceso a los resultados de la investigación, tanto en forma de publicaciones como de datos (OECD, 2015). La Unión Europea, por su parte, estableció tempranamente una política activa en favor del acceso abierto a través de la Agenda Digital para Europa y posteriormente el Programa Horizonte 2020 (Melero y Hernández San Miguel, 2014). En América Latina, la CEPAL formó parte del Proyecto Leaders Activating Research Networks (LEARN) de la League of European Research Universities (LERU) y, a partir de un relevamiento de la situación de la publicación de datos en la región, elaboró un modelo de gestión de datos de investigación y realizó tareas de promoción (Andaur, 2016b; Angelozzi, 2020). Diversos países latinoamericanos, a su vez, llevaron adelante iniciativas en la materia. Perú, Argentina y México elaboraron legislación al respecto, mientras que Chile, Brasil y Colombia llevaron adelante iniciativas de apertura de datos (Andaur, 2016a).

El movimiento *Open Data* se fundamenta en la idea de que compartir los datos redundaría en una mayor transparencia del proceso de investigación, un mayor provecho de la inversión pública y la duplicación del avance científico al reducir esfuerzos. Para esto, los datos deben ser compartidos de manera que puedan ser reutilizados a través de formatos libres, no propietarios y generalmente no textuales (Peset, Ferrer Sapena y Subirats Coll, 2011). Según la *Open Knowledge Foundation*, los datos pueden ser considerados abiertos cuando pueden ser utilizados, reutilizados y redistribuidos de manera libre por cualquier persona, con la condición de que se respete la atribución de autoría y las obras que deriven de esa reutilización se compartan en las mismas condiciones de apertura (Melero y Hernández San Miguel, 2014).

En 2014 se llevó a cabo el workshop denominado *Jointly Designing a Data Fairport*, donde una serie de especialistas interesados en el intercambio y reutilización de datos se reunieron a discutir cuáles eran los principios que ayudarían a superar los obstáculos que sufría esta práctica (Wilkinson et al., 2016). Estos principios fueron elaborados con la finalidad de guiar a los productores y quienes publican datos a que lo hagan en condiciones que garanticen que los datos puedan ser encontrados y reutilizados. Estas indicaciones no solo tienen que ver con el correcto registro y almacenado de

los datos sino también con la provisión de mecanismos que aseguren la disponibilidad a largo plazo de ellos. Se estructuran en cuatro recomendaciones y de estas se toman sus siglas: los datos deben ser encontrables (*Findable*), accesibles (*Accesible*), interoperables (*Interoperable*) y reutilizables (*Reusable*). Los principios FAIR hacen especial énfasis en que los datos puedan ser explotados tanto por máquinas como por individuos y establecen características que los recursos de datos, herramientas, vocabularios e infraestructuras deben incluir para asegurar el descubrimiento, la integración y la reutilización de los datos. Los principios están relacionados, pero son independientes y pueden ser aplicados a datos de cualquier tipo de dominio y a otros tipos de resultados de investigación (Wilkinson et al., 2016)

### 3.1. Los datos abiertos en la Argentina

En la Argentina pueden identificarse diversas iniciativas y políticas públicas en torno al Acceso Abierto de la producción científica como también a la información gubernamental. Se cuenta además con legislación en la materia, como la Ley 26.899 de Creación de Repositorios Digitales Institucionales de Acceso Abierto Propios o Compartidos (2013) para el caso de la producción científica, y la Ley 27.275 de Derecho de Acceso a la Información Pública (2016) y el Decreto 117/2016, que explicita el Plan de Apertura de Datos para los organismos públicos, en cuanto a información gubernamental.

La Ley 26.899 fue desarrollada en el marco del Ministerio de Ciencia, Tecnología e Innovación (MINCyT) a partir del trabajo conjunto de la Secretaría de Articulación Científico-Técnica de dicho ministerio y representantes de los repositorios digitales que estaban en marcha, convocados por la secretaría en el año 2009 con la idea de generar una red de repositorios. Se conformaron grupos de trabajo a partir de los cuales se señaló la necesidad de un marco legal que posibilitara el crecimiento de los contenidos de los repositorios institucionales en el país (Bongiovani y Nakano, 2011). Como resultado de este trabajo se creó el Sistema Nacional de Repositorios Digitales (SNRD) en 2011 dentro de la órbita del Programa de Grandes Instrumentos y Bases de Datos. Dos años después de la creación del SNRD, se sancionó la Ley 26.899 de 2013, que estipula que las instituciones del Sistema Nacional de Ciencia, Tecnología e Innovación (SNCTI) que reciban financiamiento del Estado nacional tienen la obligación de crear repositorios digitales institucionales donde depositar la producción científica que generan sus investigadores con fondos públicos. El alcance de esta producción no se limita a las publicaciones y tesis, sino que en su artículo 2 exige que se establezcan políticas para la gestión, depósito y preservación de los datos de investigación producidos durante la actividad científica. Establece que los repositorios deben seguir criterios de interoperabilidad, libre acceso y respeto a los derechos de autor. Finalmente, dispone plazos para el depósito de los documentos (6 meses desde su publicación) y los datos primarios (hasta 5 años luego de su recolección) y penalizaciones a su incumplimiento.

La ley indica que los repositorios que los organismos están obligados a generar pueden ser propios o compartidos con otras instituciones, y en el caso del depósito y publicación de los datos primarios, puede hacerse a través de los sistemas nacionales de grandes instrumentos y bases de datos desarrollados por el Programa de Grandes Instrumentos y Bases de Datos mencionado anteriormente. Este programa surgió conjuntamente del MINCyT y del Consejo Interinstitucional de Ciencia y Tecnología y busca realizar un uso eficiente de grandes equipamientos y de la información generada y adquirida por el Sistema Nacional de Ciencia, Tecnología e Innovación argentino. En 2017 estaba conformado por once Sistemas Nacionales de Grandes Instrumentos y cinco Sistemas Nacionales de Bases de Datos (Malvicino, 2018). Estos últimos son: Sistema Nacional de Datos Biológicos, Sistema Nacional de Datos del Mar, Sistema Nacional de Repositorios Digitales, Sistema Nacional de Datos Climáticos y Sistema

Nacional de Datos Genómicos. Cada uno de estos sistemas generaron un portal destinado a proveer acceso libre a los datos alojados en ellos, menos el de Datos Climáticos que no ha sido desarrollado y además la información sobre dicho sistema nacional fue removida del sitio web del Programa. El último sistema nacional en crearse fue el de Documentación Histórica, cuya misión es “contribuir al fortalecimiento de las capacidades nacionales de recolección, sistematización, preservación, acceso y utilización del patrimonio documental, público y privado, en sus diferentes soportes” (Argentina. MINCyT, 2017a) pero que aún no cuenta con un portal de información.

En 2017 el MINCyT a través de la Resolución 640-E/2017 creó el Programa de Datos Abiertos de Ciencia y Tecnología que tiene como objetivos:

3. <https://datos.mincyt.gob.ar/#/>
  4. <https://ri.conicet.gov.ar/>
  5. <https://proyectosinv.conicet.gov.ar/programa-ciencias-sociales/>
- » Coordinar acciones para la publicación de datos abiertos del Ministerio.
  - » Gestionar el Portal de Información en Ciencia y Tecnología argentino<sup>3</sup>.
  - » Brindar asistencia técnica a otros organismos de ciencia y tecnología.
  - » Elaborar definiciones conceptuales, requerimientos funcionales y arquitectura tecnológica.
  - » Desarrollar aplicaciones para la integración, explotación y visualización de datos en el portal.
  - » Posicionar a la Argentina como pionera en ciencia, tecnología e innovación productiva abierta. (Argentina. MINCyT, 2017b)

El Portal de Información en Ciencia y Tecnología, además de brindar información acerca del SNCTI del país, da acceso a los portales de datos científicos que gestiona este programa.

En línea al cumplimiento de la Ley 26.899, el Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), el principal organismo nacional que promueve y financia las investigaciones científicas en el país, generó su repositorio institucional Conicet Digital<sup>4</sup> para depositar las publicaciones de los investigadores que desarrollan su trabajo financiados por el Consejo y se encuentra trabajando en políticas institucionales para la gestión, almacenamiento y depósito de los datos primarios de investigación. En este sentido, desarrolló una plataforma online que permite a los investigadores desarrollar su Plan de Gestión de Datos que se encuentra en fase de prueba piloto (Actis y Carlino, 2017).

Otro proyecto del CONICET referido al tratamiento de datos de investigación es la Plataforma Interactiva de Investigación para las Ciencias Sociales (PLIICS), cuyo objetivo es “fomentar el acceso abierto a colecciones de fuentes documentales y conjuntos de datos primarios y de investigación de las Ciencias Sociales mediante la construcción de una plataforma de repositorios digitales de datos”<sup>5</sup>. En una primera etapa de este proyecto se analizaron las necesidades y las especificidades de los datos recolectados en esta área y la identificación por parte de los centros de investigación participantes de las colecciones de datos que les interesaba publicar. En una segunda etapa del programa, los centros participantes resolvieron agregar sus colecciones a repositorios universitarios o crear uno para su proyecto. Las instituciones que participaron fueron la Universidad Nacional de Córdoba, la Universidad Nacional del Centro de la Provincia de Buenos Aires y el Centro Argentino de Información Científica y Tecnológica (CAICyT) del CONICET (Pluss y Leff, 2016).

Las iniciativas desde el ámbito universitario alrededor de la gestión de los datos de investigación son escasas y se encuentran en un estadio inicial. De acuerdo con el relevamiento realizado para este trabajo, se encontró que una serie de repositorios institucionales universitarios están comenzando a incluir en sus colecciones los datos de investigación generados por sus miembros. Se destaca el caso de la Universidad

Nacional de Rosario, de la provincia de Santa Fe, que realizó una encuesta a sus investigadores docentes con el fin de explorar la producción y la gestión de datos de investigación. El objetivo de ese trabajo fue conocer las características y necesidades de los tipos de datos que se generan en la institución para poder establecer políticas de gestión institucionales adecuadas (Bongiovani y Martínez Uribe, 2014).

## 4. Relevamiento de portales y repositorios de datos de investigación en la Argentina

En este apartado se presenta el relevamiento realizado, de acuerdo a los pasos establecidos en la metodología con el fin de llevarlo a cabo de la manera más exhaustiva y rigurosa posible: 1) registros y directorios; 2) SNRD; 3) SIU-BDU; 4) búsqueda en motor de búsqueda; y 5) revisión bibliográfica. A lo largo del apartado se hace mención a las dificultades halladas, los criterios de inclusión y exclusión, y al final, se proveerá la lista completa de repositorios y portales localizadas y un breve análisis de las características del conjunto.

### 4.1. Registros y directorios de repositorios

#### 4.1.1. Registry of Research Data Repositories (Re3data)<sup>6</sup>

6. <https://www.re3data.org/>

En este registro es posible la navegación por país. La selección por Argentina arroja 7 resultados:

- » Humadoc de la Facultad de Humanidades de la Universidad Nacional de Mar del Plata (UNMDP)
- » Portal de Datos de Biodiversidad del Sistema Nacional de Datos Biológicos (SNDB)
- » Portal de Datos del Mar del Sistema Nacional de Datos del Mar (SNDM)
- » Portal de Datos Genómicos del Sistema Nacional de Datos Genómicos (SNDG)
- » Producción Académica UCC, Universidad Católica de Córdoba
- » Repositorio Institucional UCASAL, Universidad Católica de Salta
- » Servicio de Difusión de la Creación Intelectual (SEDICI), Universidad Nacional de La Plata (SEDICI-UNLP)

Al revisar el repositorio UCASAL se constató que no posee colecciones de datos en su acervo, por lo que se eliminó del listado final. De la misma manera, Humadoc fue descartado ya que los *sets* de datos que incluye consisten en fichas sobre entrevistas realizadas, pero no las entrevistas.

#### 4.1.2. OpenDOAR<sup>7</sup>

La consulta por países arrojó un total de 72 repositorios registrados de Argentina. Se ingresó a cada uno de los repositorios listados, ya que, si bien el registro de OpenDOAR incluye el tipo de materiales incluidos en los repositorios, esta información podría estar desactualizada o no haber sido completada, como se constató. La estructura de comunidades y colecciones varía mucho en cada uno de los repositorios revisados, lo que genera también diversos grados de dificultad en la detección de conjuntos de datos en ellos. Algunos repositorios ya incluyen como categoría de objeto digital al conjunto de datos (con variaciones como *set* de datos o *datasets*) de forma explícita y en otros casos se encuentran dentro de colecciones específicas objetos que consisten en datos.

7. <https://v2.sherpa.ac.uk/opensoar/>

Los repositorios extraídos de este servicio son:

- » El Abrigo, Servicio Meteorológico Nacional (SMN)
- » Filo:Digital, Repositorio Institucional de la Facultad de Filosofía y Letras de la Universidad de Buenos Aires (FFyL-UBA)
- » INTA Digital
- » Naturalis de la Facultad de Ciencias Naturales y Museo, Universidad Nacional de La Plata (FCNM-UNLP)
- » Producción Académica UCC
- » Repositorio Digital Institucional, Universidad Nacional del Comahue
- » Repositorio Digital de la UNC, Universidad Nacional de Córdoba
- » Repositorio Hipermedial, Universidad Nacional de Rosario (UNR)
- » Repositorio Institucional Digital de la Universidad Nacional de Río Negro (RID-UNRN)
- » Repositorio Institucional UCA (Universidad Católica Argentina)
- » SEDICI - UNLP
- » Suquía. Repositorio Digital del Programa de Arqueología Digital, Instituto de Antropología (IDACOR), CONICET y Museo de Antropología, Universidad Nacional de Córdoba

#### 4.1.3. Registry of Open Access Repositories (ROAR)<sup>8</sup>

8. <http://roar.eprints.org/>

La consulta realizada con el filtro por país devolvió un total de 71 repositorios registrados en ROAR. Se revisó el listado y se obtuvo que los siguientes contienen datos de investigación:

- » Filo:Digital (FFyL-UBA)
- » Naturalis (FCNM-UNLP)
- » Repositorio Digital de la UNC
- » RID-UNRN
- » Suquía

#### 4.2. Sistema Nacional de Repositorios Digitales (SNRD)

En el sitio web del SNRD se consigna que 44 repositorios están adheridos a dicho sistema. Se revisaron uno a uno para determinar la existencia de datos de investigación en sus colecciones y se determinó que los siguientes los contienen:

- » CIC Digital (Comisión de Investigaciones Científicas, Provincia de Buenos Aires)
- » El Abrigo (SMN)
- » Filo:Digital (FFyL-UBA)
- » INTA Digital
- » Naturalis (FCNM-UNLP)
- » Producción Académica UCC
- » Repositorio Digital de la UNC
- » Repositorio Digital Institucional, Universidad Nacional del Comahue
- » RID-UNRN
- » Repositorio Institucional UCA
- » SEDICI - UNLP

#### 4.3. SIU-BDU<sub>2</sub>

En primer lugar, se realizó una búsqueda por tipo de formato. El sistema ofrece dos tipos de formatos que se adecuan a nuestra búsqueda: conjunto de datos y *set* de datos.

La búsqueda por conjunto de datos arrojó 1886 resultados que eran provenientes del SEDICI (UNLP), Suquía (IDACOR-CONICET-UNC), El Abrigo (SMN) y del

Repositorio Institucional de la UCA. En cambio, la búsqueda por el tipo de formato *set* de datos arrojó 1737 resultados que provenían de distintas universidades:

- » Filo:Digital (FFyL-UBA)
- » Humadoc (UNMDP)
- » Repositorio Digital de la UNC
- » Repositorio Hipermedial (UNR)
- » Repositorio Institucional SEGEMAR. Servicio Geológico Minero Argentino, Secretaría de Minería, Ministerio de Desarrollo Productivo
- » Rpsico. Facultad de Psicología de la Universidad Nacional de Mar del Plata,
- » Suquía

Como en este listado aparecían repositorios que habían sido desechados en la revisión de los directorios y el SNRD, se procedió a observar en qué consistían los objetos digitales indicados como *set* de datos. El único objeto recuperado del Repositorio de la Facultad de Psicología de la Universidad Nacional de Mar del Plata era un “Boletín de Relaciones Internacionales” pero no se pudo acceder al sitio por encontrarse fuera de servicio. En el caso del Repositorio Institucional Servicio Geológico Minero de la Argentina, se trataba de hojas geológicas que, si bien contienen datos geológicos y son la base de la investigación en esa disciplina, las alojadas en el repositorio consisten en un mapa y su informe explicativo y están en formato PDF. Veremos en el siguiente apartado que los mismos datos pueden estar en formatos reutilizables en un sistema de información geográfica y ser considerados datos primarios. El repositorio Humadoc fue descartado por las razones expuestas en la sección de directorios.

#### 4.4. Búsqueda de información y revisión bibliográfica

Se realizaron búsquedas en el motor de búsqueda más popular, Google, en una pestaña privada del navegador para que los resultados no fueran influenciados por nuestro historial de búsqueda. Google devuelve millones de resultados a las búsquedas realizadas por lo que se revisaron las primeras 10 páginas de resultados de cada una de las estrategias de búsqueda planteadas. La búsqueda por *portal datos abiertos investigación argentina* arroja 8 millones de resultados, de los cuales el primero es el Portal de Información en Ciencia y Tecnología. La búsqueda por *repositorio datos abiertos investigación argentina* devolvió más de 4 millones de resultados, pero se observó una relevancia menor en relación con la anterior búsqueda. No solo se revisaron sitios específicos que remiten a portales de datos y repositorios, sino que también se observaron algunas noticias que podían referir a lanzamientos o proyectos en curso en busca de herramientas que sirvan a esta investigación.

Los recursos de información hallados en esta indagación son:

- » Portal de Información en Ciencia y Tecnología:
  - Portal de Datos Genómicos (SNDG)
  - Portal de Datos de Biodiversidad (SNDB)
  - Portal de Datos del Mar (SNDM)
  - Catálogo de Imágenes y Productos de la Comisión Nacional de Actividades Espaciales
- » Datos abiertos del Ministerio de Salud
- » Portal de Datos Económicos del Ministerio de Economía
- » Banco Central de la República Argentina
- » Datos Agricultura, Ganadería y Pesca del Ministerio de Agricultura, Ganadería y Pesca

- » Suelos de Córdoba: Portal de datos abiertos de suelos de la Provincia de Córdoba. Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba
- » INCIHUSA Digital. Instituto de Ciencias Humanas, Sociales y Ambientales, CONICET
- » Biblioteca de datos abiertos sobre educación de la Fundación Luminis

#### **4.5. Portales identificados en otras instancias de búsqueda y relevamiento bibliográfico**

En este apartado se da cuenta de repositorios o portales de datos que fueron identificados o se tomó conocimiento de ellos a través de fuentes documentales, redes sociales o eventos profesionales, que no fueron localizados en las instancias anteriores, y que fueron evaluados para su inclusión en el relevamiento.

- » Poblaciones. Plataforma abierta de datos espaciales de la Argentina desarrollada por el Observatorio de la Deuda Social Argentina de la Universidad Católica Argentina (ODSA-UCA) y CONICET
- » Ministerio de Ambiente y Desarrollo Sostenible
- » BiodAr: Biodiversidad de Insectos de la Argentina y Uruguay. CONICET y UNLP
- » Sistema de Información de Biodiversidad. Administración de Parques Nacionales de Argentina
- » Sistema de Información Geológica Ambiental Minera (SIGAM). SEGEMAR, Ministerio de Desarrollo Productivo
- » Repositorio Digital Archivo DILA. Centro Argentino de Información Científica y Tecnológica, CONICET
- » Archivo Digital Comunitario en el Repositorio Digital de Ciencias Humanas de la Universidad Nacional del Centro de la Provincia de Buenos Aires (UNICEN)
- » Infraestructura de Datos Espaciales del Instituto de Geografía, Historia y Ciencias Sociales (IGEHCs), CONICET-UNICEN y Facultad de Ciencias Humanas, UNICEN
- » IDERA: Infraestructura de Datos Espaciales de la República Argentina
- » Observatorio Medioambiental La Plata. Universidad Nacional de La Plata, CONICET y Comisión de Investigaciones Científicas de la Provincia de Buenos Aires

#### **4.6. Listado definitivo de repositorios a analizar**

Luego del análisis de cada uno de los portales hallados, se decide eliminar de la muestra a aquellos que no respondan a la definición de nuestro objeto de estudio: repositorios y portales que reúnan datos producidos en el desarrollo de la investigación científica. Los portales de datos abiertos de los Ministerios de Ambiente y Desarrollo Sostenible, Agricultura, Ganadería y Pesca, Salud y Hacienda responden a la definición de portal de transparencia, cuyo objetivo es publicar los datos generados por las reparticiones estatales durante la ejecución de sus funciones, no en el contexto de investigaciones científicas. IDERA es una plataforma que contiene datos georreferenciados provenientes de diversos organismos de gobierno, entre los que se cuentan los mencionados anteriormente. El sitio del Banco Central de la República Argentina tampoco se ajusta a la descripción de nuestro objeto de estudio al igual que el portal de Datos Abiertos sobre Educación de la Fundación Luminis. Este último caso se compone mayormente de datos estadísticos tomados del Instituto Nacional de Estadística y Censos de la República Argentina (INDEC) y del Ministerio de Educación de Argentina.

De los repositorios enmarcados en la iniciativa PLIICS del CONICET, no tomaremos el caso del Archivo Digital Comunitario en el Repositorio Digital de Ciencias Humanas

de la Universidad Nacional del Centro de la Provincia de Buenos Aires, ya que como su nombre lo indica, contiene material de archivo: fotografías mayormente.

Finalmente, retiramos de la selección a Poblaciones: plataforma abierta de datos espaciales de la Argentina, ya que esta herramienta no tiene como principal objetivo alojar datos de investigación, sino que su principal función es generar visualizaciones de datos de forma georreferenciada.

Tabla 1: Listado definitivo de repositorios y portales

Repositorio	Institución	URL
BiodAr: Biodiversidad de Insectos de la Argentina y Uruguay	CONICET y UNLP	<a href="https://biodar.unlp.edu.ar/es/catalogues/">https://biodar.unlp.edu.ar/es/catalogues/</a>
Catálogo de Imágenes y Productos	Comisión Nacional de Actividades Espaciales, MINCyT	<a href="https://catalogos.conae.gov.ar/catalogo/catalogo.html">https://catalogos.conae.gov.ar/catalogo/catalogo.html</a>
CIC Digital	Comisión de Investigaciones Científicas de la Provincia de Buenos Aires	<a href="https://digital.cic.gba.gob.ar/">https://digital.cic.gba.gob.ar/</a>
El Abrigo (SMN)	Servicio Meteorológico Nacional	<a href="http://repositorio.smn.gob.ar/">http://repositorio.smn.gob.ar/</a>
Filo:Digital	Facultad de Filosofía y Letras, Universidad de Buenos Aires	<a href="http://repositorio.filo.uba.ar/xmlui/">http://repositorio.filo.uba.ar/xmlui/</a>
INCIHUSA Digital	Instituto de Ciencias Humanas, Sociales y Ambientales, CONICET	<a href="http://incihusa.mendoza-conicet.gob.ar/id/">http://incihusa.mendoza-conicet.gob.ar/id/</a>
Infraestructura de Datos Espaciales (UNICEN)	Facultad de Ciencias Humanas, UNICEN, IGEHCS, CONICET	<a href="http://ide.fch.unicen.edu.ar/">http://ide.fch.unicen.edu.ar/</a>
INTA Digital	Instituto Nacional de Tecnología Agropecuaria	<a href="https://repositorio.inta.gob.ar/">https://repositorio.inta.gob.ar/</a>
Naturalis (FCNM-UNLP)	Facultad de Ciencias Naturales y Museo, UNLP	<a href="http://naturalis.fcnym.unlp.edu.ar">http://naturalis.fcnym.unlp.edu.ar</a>
Observatorio Medioambiental La Plata	UNLP, CONICET y Comisión de Investigaciones Científicas de la Provincia de Buenos Aires	<a href="http://omlp.sedici.unlp.edu.ar/">http://omlp.sedici.unlp.edu.ar/</a>
Portal de Datos de Biodiversidad (SNDB)	MINCyT	<a href="https://datos.sndb.mincyt.gob.ar/ala-hub/search">https://datos.sndb.mincyt.gob.ar/ala-hub/search</a>
Portal de Datos del Mar (SNDM)	MINCyT	<a href="http://portal.mincyt.gob.ar/portal/portal/sndm/home">http://portal.mincyt.gob.ar/portal/portal/sndm/home</a>
Portal de Datos Genómicos (SNDG)	MINCyT	<a href="https://datos.sndg.mincyt.gob.ar/">https://datos.sndg.mincyt.gob.ar/</a>
Producción Académica UCC	Universidad Católica de Córdoba	<a href="http://pa.bibdigital.uccor.edu.ar/">http://pa.bibdigital.uccor.edu.ar/</a>

Repositorio Digital Archivo DILA	Centro Argentino de Información Científica y Tecnológica, CONICET	<a href="http://www.caicyt-conicet.gov.ar/dila/">http://www.caicyt-conicet.gov.ar/dila/</a>
Repositorio Digital Institucional	Universidad Nacional del Comahue	<a href="http://rdi.uncoma.edu.ar/">http://rdi.uncoma.edu.ar/</a>
Repositorio Digital de la UNC	Universidad Nacional de Córdoba	<a href="https://rdu.unc.edu.ar">https://rdu.unc.edu.ar</a>
Repositorio Hipermedial (UNR)	Universidad Nacional de Rosario	<a href="https://rephip.unr.edu.ar/">https://rephip.unr.edu.ar/</a>
Repositorio Institucional UCA	Universidad Católica Argentina	<a href="https://repositorio.uca.edu.ar">https://repositorio.uca.edu.ar</a>
RID-UNRN	Universidad Nacional de Río Negro	<a href="https://rid.unrn.edu.ar/jspui/">https://rid.unrn.edu.ar/jspui/</a>
SEDICI-UNLP	Universidad Nacional de La Plata	<a href="http://sedici.unlp.edu.ar">http://sedici.unlp.edu.ar</a>
Sistema de Información de Biodiversidad	Administración de Parques Nacionales de Argentina	<a href="https://sib.gob.ar/portada">https://sib.gob.ar/portada</a>
SIGAM	SEGEMAR, Ministerio de Desarrollo Productivo	<a href="https://sigam.segemar.gov.ar/">https://sigam.segemar.gov.ar/</a>
Suelos de Córdoba: Portal de datos abiertos de suelos de la Provincia de Córdoba	Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba	<a href="http://sueloscordoba-agrounc.opendata.arcgis.com/">http://sueloscordoba-agrounc.opendata.arcgis.com/</a>
Suquia	IDACOR-CONICET, Museo de Antropología, Universidad Nacional de Córdoba	<a href="https://suquia.ffyh.unc.edu.ar/">https://suquia.ffyh.unc.edu.ar/</a>

El listado queda compuesto entonces de 25 repositorios y portales, de los cuales el 36%, es decir 9, están destinados exclusivamente a la publicación de datos de investigación; mientras que el 64% (16) publican tanto documentos como datos. Dentro de los últimos, se cuentan repositorios con colecciones muy pequeñas de datos, 8 de ellos tienen entre 1 y 4 conjuntos de datos (Gráfico 1).

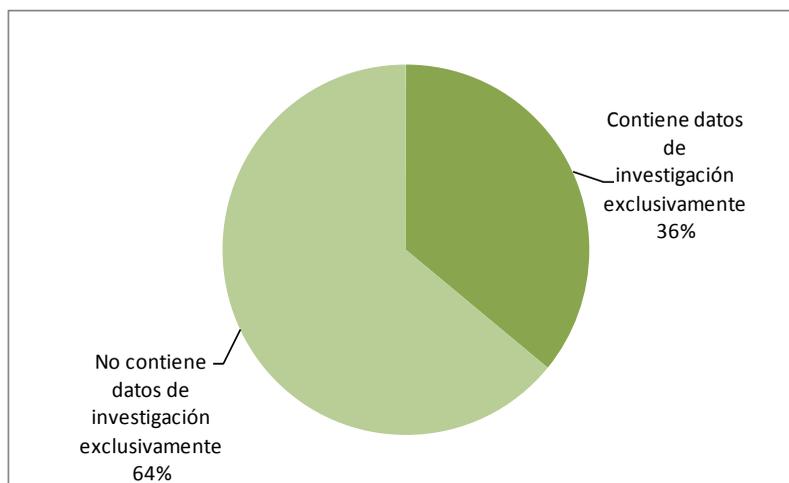
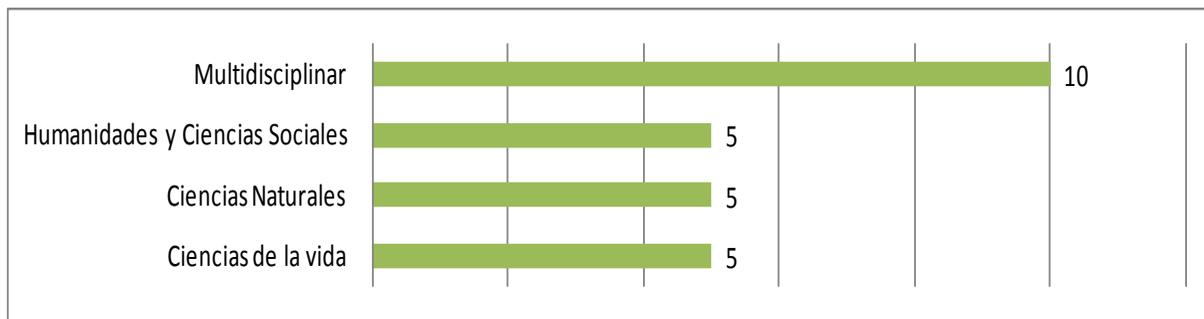


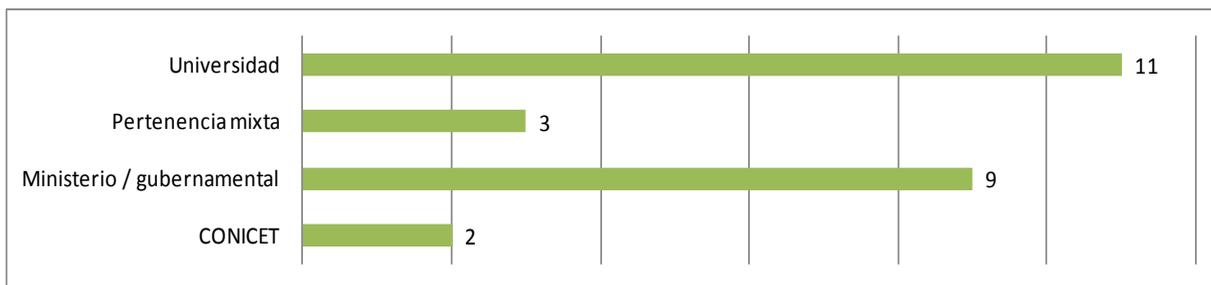
Gráfico 1: Contenido de los repositorios y portales

Gráfico 2: Distribución de repositorios según áreas temáticas de Re3data



Fuente: elaboración propia

Gráfico 3: Distribución de repositorios según tipo de institución a la que pertenecen



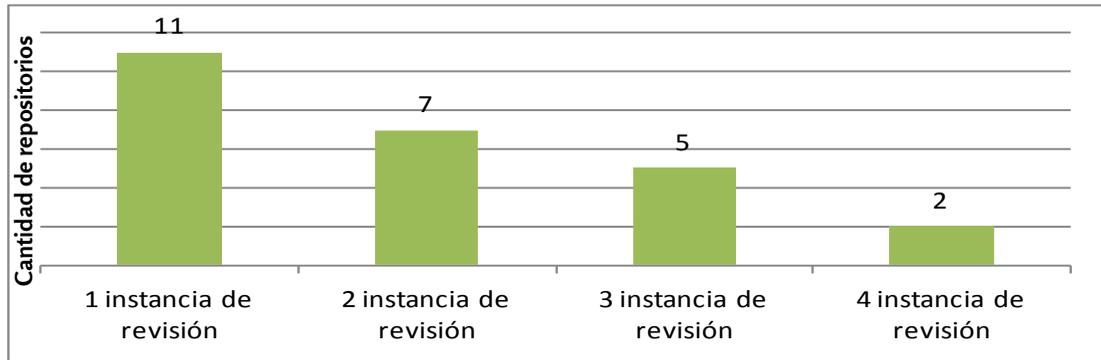
Fuente: elaboración propia

Con el fin de visualizar la distribución por áreas temáticas de estos repositorios, los clasificamos tomando las disciplinas amplias definidas por Re3data. Pudimos observar que las grandes áreas disciplinares de las Humanidades y Ciencias Sociales, las Ciencias Naturales (Geología, Astronomía, Oceanografía) y las Ciencias de la Vida (Biología, Genética) están representadas por 5 repositorios cada una, lo que significa 20% del total por área. Finalmente, el 40% restante (10 repositorios) se corresponde con la categoría de plataformas multidisciplinares, ya que contienen información de dos o más de las categorías anteriores (Gráfico 2).

Observamos que el 40% de los repositorios relevados pertenecen a universidades (10 en total). Los siguen en cantidad los pertenecientes a ministerios públicos u otras agencias gubernamentales que ascienden al 36% del total, con 9 servicios. Dos repositorios, el 8% del total, pertenecen al Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET). Finalmente, un 16% -4 en total- tienen una pertenencia mixta: CONICET y universidad en tres casos, y en el último, CONICET, universidad y agencia provincial (Comisión de Investigaciones Científicas de la Provincia de Buenos Aires) (Gráfico 3).

Para concluir este breve análisis del conjunto de repositorios y portales de datos hallados, se analizó la presencia de cada uno de ellos en las distintas instancias del relevamiento. Se encontró que el 41% de los recursos solo se halló en una instancia, un 26% se encontró en 2 de los pasos indicados en la metodología, un 19% en 3 instancias, mientras que solo 2 estaban presentes en 4 de las instancias planteadas para el relevamiento (Gráfico 4).

Gráfico 4: Difusión y visibilidad de los repositorios y portales



Fuente: elaboración propia

Estos guarismos son indicadores del nivel de difusión y visibilidad de estos recursos de información y son producto de las políticas de posicionamiento de los repositorios y portales por parte de las instituciones responsables.

## 5. Conclusiones

A partir del relevamiento realizado, se observa un crecimiento de casi un 50% en la cantidad de repositorios y portales que incluyen datos de investigación hallados en el año 2021 con respecto a los analizados en el Trabajo de Fin de Máster realizado en 2020 (Indart, 2020): de 17 a 25 en total. A pesar de este aumento considerable, la cifra resulta baja al tener en cuenta el marco normativo vigente para todos los organismos del sistema científico nacional.

Un aspecto importante a señalar es la dificultad que conlleva realizar este tipo de relevamiento debido a que muchos repositorios no se encuentran registrados adecuadamente en los directorios existentes como Re3data, OpenDOAR y ROAR. Si bien hay un alto grado de solapamiento entre los dos últimos recursos, hay numerosos repositorios que solo se encuentran en uno de ellos. Asimismo, la calidad de la información sobre los repositorios que contienen estos directorios es muy heterogénea, muchas veces la información no está actualizada y, específicamente, la descripción del tipo de documento que contienen los repositorios no es exhaustiva. En el caso de OpenDOAR, si se busca por conjunto de datos y se restringe por Argentina, se obtienen solo 3 resultados; mientras que la misma búsqueda en ROAR no arroja ninguno. A esta dificultad se suma la heterogeneidad de los datos de acuerdo a las diferentes disciplinas, y el desafío que conlleva reconocer y valorar qué es un dato primario de investigación en cada marco disciplinario para una persona externa a ellos.

El 64% de los repositorios y portales localizados contienen diversos tipos documentales y en muchos de ellos la presencia de datos primarios es marginal. Ante esto, es posible preguntarse por el grado de adecuación del tipo de plataforma utilizada para la identificación y reutilización de datos. El presente estudio se complementa con el análisis de los repositorios y portales identificados a través de este relevamiento para evaluar su implementación de acuerdo a criterios de calidad validados por la comunidad especializada que, si bien no responde de forma acabada la pregunta por la adecuación, ayudará a valorar los esfuerzos realizados. En un futuro se podrá continuar este trabajo, además, con el relevamiento y análisis de portales de datos gubernamentales ya que muchos de ellos resultan de relevancia para realizar investigaciones científicas y completan el movimiento por los datos abiertos.

## Referencias bibliográficas

- » Actis, Guillermina y Lorena Carlino. 2017. Plan de Gestión de Datos en CO-NICET: análisis, experiencia y desafíos. En *Conferencia Internacional sobre Bibliotecas y Repositorios Digitales de América Latina (7a: 2 al 4 octubre 2017: La Plata)*. VII Conferencia Internacional BIREDIAL-ISTEC'17: memoria final. La Plata: Universidad Nacional de La Plata. p. 215-231. <<http://sedici.unlp.edu.ar/handle/10915/63588>> [Consulta: 29 julio 2021].
- » Aleixandre Benavent, Rafael, Antonia Ferrer Sapena y Fernanda Peset. 2021. Compartir los recursos útiles para la investigación: datos abiertos (open data). En *Educación Médica*. Vol. 2, sup. 3, 208-215. <<https://doi.org/10.1016/j.edu-med.2019.07.004>>
- » Andaur, Gabriela. 2016a. *Panorama de la gestión de datos de investigación en América Latina y El Caribe*. <<http://learn-rdm.eu/es/gestion-de-datos-de-investigacion-en-america-latina/>> [Consulta: 29 julio 2021].
- » Andaur, Gabriela. 2016b. *Preparando el camino hacia la gestión de datos de investigación en América Latina y el Caribe: el rol de la CEPAL en el Proyecto LEARN*. <<http://learn-rdm.eu/es/preparando-el-camino-hacia-la-gestion-de-datos-de-investigacion-en-america-latina/>> [Consulta: 29 julio 2021].
- » Angelozzi, Silvina. 2020. La gestión de datos de investigación en abierto: introducción al rol emergente para las bibliotecas universitarias y científicas argentinas. En *Palabra Clave (La Plata)*. Vol. 9, no. 2, e091. <<https://doi.org/10.24215/18539912e091>>
- » Argentina. Ministerio de Ciencia, Tecnología e Innovación Productiva (MINCyT). 2017a. *Creación Sistema Nacional de Documentación Histórica (RESOL-2017-714-APN-MCT)*. <[https://back.argentina.gob.ar/sites/default/files/resol-2017-714-apn-mct\\_creacion\\_sndh.pdf](https://back.argentina.gob.ar/sites/default/files/resol-2017-714-apn-mct_creacion_sndh.pdf)> [Consulta: 4 noviembre 2021].
- » Argentina. Ministerio de Ciencia, Tecnología e Innovación Productiva (MINCyT). 2017b. *Programa de datos abiertos en ciencia y tecnología (Resolución 640-E/2017)* <<https://www.boletinoficial.gob.ar/detalleAviso/primera/167874/20170724?busqueda=1>> [Consulta: 4 noviembre 2021].
- » Bongiovani, Paola y Luis Martínez Uribe. 2014. Necesidades de gestión de datos científicos en Argentina. El caso de la Universidad Nacional de Rosario. En *Conferência Internacional Acesso Aberto, Preservação Digital, Interoperabilidade, Visibilidade e Dados Científicos Biredial-ISTEC (4a: 2014: Porto Alegre)*. Porto Alegre: Universidade Federal do Rio Grande do Sul. p. 104-121. <<http://hdl.handle.net/2133/4384>> [Consulta: 29 julio 2021].
- » Bongiovani, Paola y Silvia Nakano. 2011. Acceso abierto en Argentina: la experiencia de articulación y coordinación institucional de los repositorios digitales en ciencia y tecnología. En *E-colabora*. Vol 1, no. 2, 163-179.
- » Borgman, Christine L. 2012. The conundrum of sharing research data. En *Journal of the American Society for Information Science and Technology*. Vol. 63, no. 6, 1059-1078. <<https://doi.org/10.1002/asi.22634>>
- » Borgman, Christine L. 2015. *Big data, little data, no data: Scholarship in the networked world*. Cambridge: MIT Press.

- » Indart, Camila. 2020. *Datos abiertos de investigación en Argentina: análisis de su implementación en portales y repositorios*. Salamanca: Universidad de Salamanca, 2020. 80, liv p. Trabajo de fin de Máster. <<http://hdl.handle.net/10366/147099>> [Consulta: 22 octubre 2021].
- » Malvicino, Facundo. 2018. *Evaluación del Programa de Grandes Instrumentos y Bases de Datos*. CIECTI. <<http://www.ciecti.org.ar/publicaciones/it11-evaluacion-del-programa-de-grandes-instrumentos-y-bases-de-datos/>> [Consulta: 29 julio 2021].
- » Melero, Remedios y José Hernández San Miguel. 2014. Acceso abierto a los datos de investigación, una vía hacia la colaboración científica. En *Revista Española de Documentación Científica*. Vol 37, no. 4, eo66. <<https://doi.org/10.3989/redc.2014.4.1154>>
- » National Science Board. 2005. *Long-lived digital data collections: Enabling Research and Education in the 21st Century*. <<http://www.nsf.gov/pubs/2005/nsbo540/>> [Consulta: 29 julio 2021].
- » Organisation for Economic Cooperation and Development (OECD). 2007. *OECD Principles and Guidelines for Access to Research Data from Public Funding*. Paris: OECD Publishing. <<https://www.oecd.org/sti/inno/38500813.pdf>> [Consulta: 29 julio 2021].
- » Organisation for Economic Cooperation and Development (OECD). 2015. *Making Open Science a Reality*. Paris: OECD Publishing. <<http://dx.doi.org/10.1787/sjrs2f963zs1-en>>
- » Peset, Fernanda, Antonia Ferrer Sapena e Imma Subirats Coll. 2011. Open data y Linked open data: su impacto en el área de bibliotecas y documentación. En *El Profesional de la Información*. Vol. 20, no. 2, 165-173. <<https://doi.org/10.3145/epi.2011.mar.06>>
- » Pluss, Ricardo y Laura Leff. 2016. *Lineamientos para la gestión de datos científicos nacionales: la experiencia de la iniciativa PLIICS en Argentina*. Trabajo presentado en el II Congreso Argentino de Estudios Sociales de la Ciencia y la Tecnología, realizado en Bariloche del 30 de noviembre al 2 de diciembre de 2016.
- » Torres Salinas, Daniel, Nicolás Robinson García y Álvaro Cabezas Clavijo. 2012. Compartir los datos de investigación en ciencia: introducción al *data sharing*. En *El profesional de la Información*. Vol. 21, no. 2, 173-184. <<http://dx.doi.org/10.3145/epi.2012.mar.08>>
- » Wilkinson, Mark, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. En *Scientific Data*. No. 3, 160018. <<https://doi.org/10.1038/sdata.2016.18>>