

Retos y oportunidades de la sistematización y reutilización de los datos de investigación

EDER ÁVILA BARRIENTOS

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

eder@iibi.unam.mx

Introducción

Desde hace más de una década, los datos de investigación se encuentran en el foco de atención de las comunidades académicas y científicas, esto gracias a las nuevas características para el desarrollo de investigación basada en el descubrimiento intensivo en datos que se enmarcan en un contexto de latente apertura a la documentación, contenidos y datos que son generados como parte de las actividades científicas.

Proyectos como R3Data, Dryad y Zenodo, ponen de manifiesto la generación de sistemas de información con nuevas características, tanto estructurales como conceptuales, para gestionar a los datos de investigación que son generados en diversos espacios disciplinarios. Estos sistemas son significativos no solo por el nivel de tecnología que emplean, sino por la posibilidad de realizar procesos de interacción con los usuarios que los utilizan para satisfacer una determinada demanda informativa. Pues “los datos juegan un papel importante en la gestión de desastres, el pronóstico del tiempo, las decisiones de política gubernamental y para lograr los objetivos de desarrollo sostenible. Los avances en la tecnología digital y los instrumentos de observación han llevado a un crecimiento exponencial de los datos (UNESCO, 2022: párr. 1)”.



Fig. 1 Ciclo de vida de los datos de investigación. Fuente: <https://library.ethz.ch/en/publishing-and-archiving/data-management-and-policies/research-data-management/research-data-life-cycle.html>

De esta manera, cuando una comunidad de investigadores trabaja a diario con una amplia gama de datos de investigación. En el curso de un proyecto, estos datos pasan por varias fases del llamado ciclo de vida de datos de investigación (véase figura 1) hasta que terminan en su publicación, se archivan a largo plazo o se eliminan de acuerdo con el plan de gestión de sus datos. Sin embargo, en las instituciones dedicadas a la investigación existen una serie de interrogantes relacionadas con las ventajas y desventajas de sistematizar y reutilizar los datos que son generados por sus comunidades académicas y científicas. Pues un cambio de paradigma relacionado con la forma de publicar y visibilizar el conocimiento académico y científico pone de manifiesto una nueva vía para el intercambio y conexión de saberes en el entorno digital. Bajo esta premisa, el propósito de este trabajo consiste en analizar los retos y oportunidades de la sistematización y reutilización de datos de investigación, tomando como referente las variables que se manifiestan en el contexto actual.

Datos de investigación

El concepto de datos de investigación puede resultar bastante amplio y complejo. Esto debido a la amplia variedad y naturaleza de los datos que están disponibles en los múltiples dominios de la realidad. El problema de la definición del concepto de los datos resulta abstracto en los contextos de

la teoría de la ciencia de la información, dando lugar a múltiples enfoques de la percepción y uso de los datos. En este sentido, Floridi (2016, p.4) manifiesta que “los datos son cúmulos agrupados de manera adecuada, siguiendo reglas (sintaxis) que gobiernan un sistema escogido, el código o el lenguaje que se está analizando”.

Es decir, los datos forman parte de un determinado contexto y están codificados mediante un sistema y lenguaje particular. Además, “el suministro de datos avanza rápidamente, mientras que la capacidad de utilizarlos se retrasa mucho” (Lohr 2015, 35) pues la liberación de grandes cantidades de datos por parte de sus creadores pone de manifiesto problemáticas relacionadas con su correcto manejo y organización. La revolución de los datos es un fenómeno caracterizado por el constante uso y la explotación de datos en diferentes ámbitos de la actividad humana. De ello deriva la aparición de principios como el dataísmo que expone un culto a los datos basado en el uso de tecnología computacional y algorítmica para acelerar su procesamiento y aprovechamiento.

Este fenómeno lleva a considerar que los datos pueden tener diversa tipología, desde aquella que va de lo textual, numérico y alfanumérico, hasta aquella que responde a una categoría más amplia relacionada con su carácter cuantitativo o cualitativo. Esta última categoría es la que con más frecuencia se utiliza al momento de distinguir a los datos de investigación desde su perspectiva holística.

A su vez, el intercambio de datos presenta como tendencia reciente que los datos de investigación se encuentren liberados para ser usados por terceros, siguiendo el modelo open data y de open access a las publicaciones. El cuarto paradigma de la ciencia basado en el descubrimiento intensivo en datos supone la explotación de los datos de investigación para responder a preguntas y problemas de amplia complejidad en cortos periodos de tiempo, esto apoyado en el uso de las supercomputadoras, capaces de procesar cantidades enormes de datos (big data).

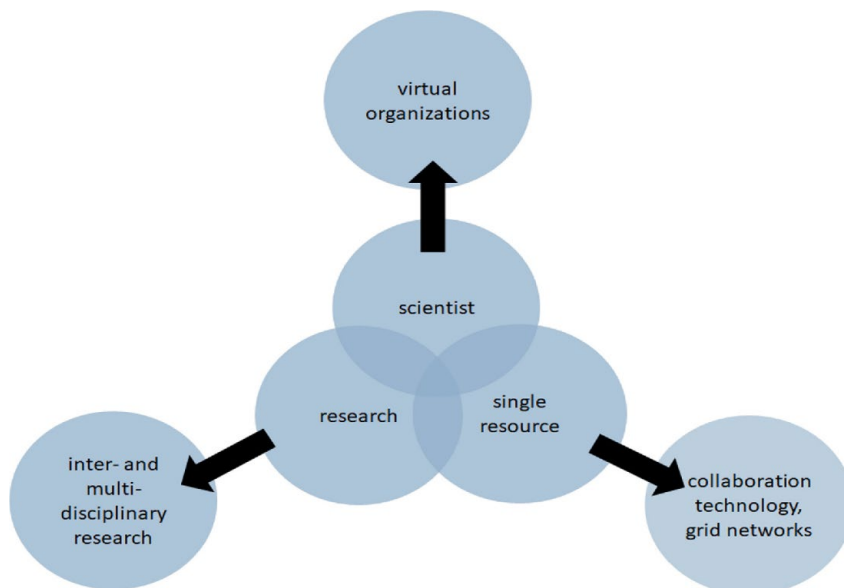


Fig 2. E-science framework including collaboration.

Fuente: Koschtial, Claudia. 2021. Disponible en: https://link.springer.com/chapter/10.1007/978-3-030-66262-2_1/figures/2

Bajo esta premisa, los datos de investigación son piezas de información que son generados en los diferentes campos disciplinarios. Pueden ser de carácter cuantitativo o cualitativo y son creados u obtenidos mediante la aplicación de diferentes metodologías e instrumentos. La accesibilidad a los datos de investigación puede ser abierta o cerrada.

A su vez, se estima que para fomentar la utilización y el aprovechamiento de este tipo de datos en la generación de nuevos conocimientos, sobre todo de manera rápida, en la actualidad se requiere que se encuentren liberados de manera abierta. Esta situación conlleva la aparición de una corriente filosófica relacionada con la apertura a los datos de investigación para el desarrollo de la e-ciencia (véase figura 2).

En tiempos actuales de pandemia mundial, las comunidades virtuales de científicos e investigadores han contribuido en gran manera en la consolidación de un entorno virtual de investigación, en donde el uso de los datos y la colaboración multi e interdisciplinaria son trascendentales para el desarrollo de la investigación de frontera.

Diversos grupos y comunidades de investigación han optado por abrir sus datos de investigación con la intención de que puedan ser accesibles y

reutilizados en el ambiente digital. Por ejemplo, el proyecto Zenodo (<https://zenodo.org/>) es un repositorio de datos que almacena, organiza, da acceso y permite descargar a conjuntos de datos que son generados por comunidades científicas y de investigación multidisciplinarias, con la peculiaridad de poder citar y dar reconocimiento a los autores y creadores de los datos de investigación que han sido depositados en su plataforma.

Dryad (<https://datadryad.org/stash>) es otro ejemplo de la utilidad y aplicación de los repositorios de datos para dar apertura a los *datasets* que han sido desarrollados por comunidades académicas. En este repositorio, puede observarse la integración de diversos componentes tecnológicos, relacionados con la aplicación de identificadores persistentes para establecer relaciones entre los creadores de los conjuntos de datos y sus instituciones. Por ejemplo, Dryad permite utilizar identificadores como ORCID y DOI para establecer un mecanismo de interacción entre los datos de investigación, su contenido y los recursos de información con los que pueden vincularse.

A su vez, el proyecto R3Data (<https://www.re3data.org/>) se trata de un registro de repositorios de datos de investigación que esta enriquecido con un esquema de metadatos desarrollado específicamente para describir y representar a conjuntos de datos de investigación. El esquema de metadatos Datacite (<https://datacite.org/>) proporciona 19 elementos descriptivos que son utilizados para registrar y representar los atributos informativos e intelectuales de los datos de investigación.

La extensibilidad del esquema Datacite ha llevado a desarrollar una iniciativa integral denominada DataciteCommons (<https://commons.datacite.org/>), la cual sirve como una interfaz de búsqueda web basada en grafos. Los grafos son formados mediante la recopilación de recursos de información académicos, como publicaciones, conjuntos de datos, personas y organizaciones de investigación, y sus respectivas conexiones. Dicha interfaz utiliza identificadores persistentes y un lenguaje de programación basado en grafos (GraphQL), los cuales son construidos mediante el análisis de metadatos proporcionados por DataCite, Crossref, ORCID entre otras plataformas de identificadores persistentes.

Como puede apreciarse, el uso de los metadatos para describir y dotar de accesibilidad a los datos de investigación es crucial. Pues los metadatos, fomentan el desarrollo de vinculaciones entre dichos datos, sus creadores e instituciones. Bajo esta premisa, “uno de los principales argumentos para la identificación y documentación de las buenas prácticas y formatos consistentes de metadatos para los servicios de datos de investigación es crear una

base sólida sobre la cual se pueda garantizar la interoperabilidad semántica y temática de dichos datos (Farnel y Shiri, 2014: p. 75)”. Por lo tanto, la apertura de los datos de investigación y su correcta descripción y representación mediante el uso de metadatos puede fomentar acciones para su latente reutilización en un entorno digital cada vez más proliferado de datos con diversas características y complejidades referentes a su manejo.

Concepto de reutilización

“Si he logrado ver más lejos, ha sido porque he subido a hombros de gigantes”. La frase que Isaac Newton enviaría a Robert Hooke en una carta el 5 de febrero de 1675 resume de manera puntual el carácter universal del desarrollo del conocimiento científico. Pues a través de la historia el progreso de la ciencia ha tenido un sentido colectivo que permite contribuir a la búsqueda de respuestas ante las problemáticas y fenómenos que se presentan en la realidad.

En el contexto tecno-digital actual y con el auge de la investigación de carácter interdisciplinario, es frecuente el uso de datos de investigación para refutar o comprobar teorías que aborden un determinado problema o encontrar respuestas a fenómenos de múltiples aristas. Por ejemplo, “los datos de un estudio anterior se pueden volver a analizar para validar, verificar o confirmar la investigación primigenia en un estudio de reproducibilidad, donde se vuelve a hacer la misma pregunta utilizando los mismos datos y métodos de análisis (Borgman, 2015)”. El carácter de reproducibilidad de los datos se apoya en el principio de citación, que permite dar crédito a la idea original planteada en una investigación, respetando la propiedad intelectual de los investigadores.

En este sentido, la reutilización de los datos de investigación va más allá de compartir datos y de divulgarlos en un sistema *ex profeso*, se trata de tomar a dichos datos y sujetarlos a un análisis tomando en cuenta a las metodologías e instrumentos de la investigación primaria de donde fueron recabados. Por ende, los datos de investigación deberán contar con la documentación correspondiente que explique su alcance y utilidad en un proceso formal de investigación. No obstante, “reutilizar un solo conjunto de datos en su forma original es difícil, incluso si se dispone de la documentación y las herramientas adecuadas, ya que se debe entender mucho acerca de por qué se recopilaban los datos y por qué se tomaron varias decisiones sobre la recopilación, limpieza y análisis de datos (Pasqueto, Randles y Borgman, 2017: p. 4)”.

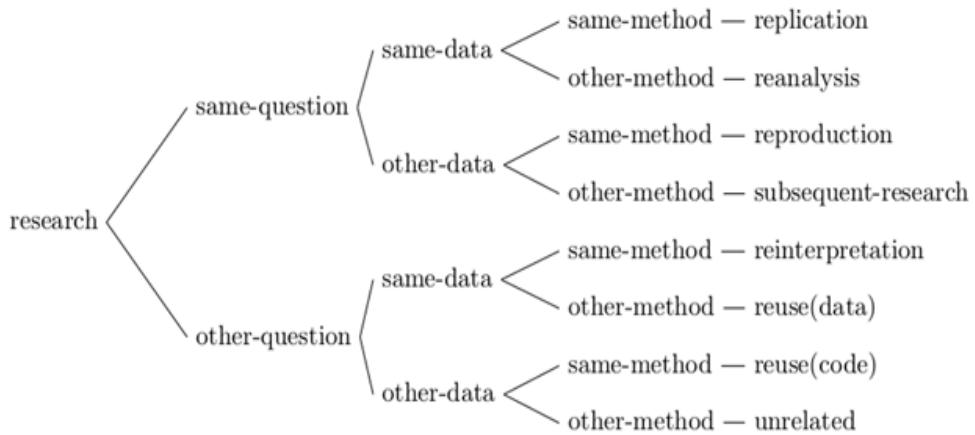


Fig 3. The concept of reusable research, adapted from Schöch (2017). Fuente: Schöch, C. 2017. Wiederholende Forschung in den digitalen Geisteswissenschaften: <https://christofs.github.io/wiederholende-forschung-dhd/>.

En la figura 3, puede apreciarse el proceso de reutilización dentro del contexto de la investigación. Destaca la presencia del concepto de reutilización al momento de establecerse en el marco común de una investigación. En donde, una determinada investigación puede abordar otra pregunta, utilizar los mismos datos de un estudio previo y a través de la implementación de otro método, se puede desarrollar la reutilización de los datos en un sentido de abstracción lógica.

Por lo tanto, los datos desarrollados en una investigación original y primigenia pueden ser reutilizados mediante la aplicación de otro método diferente al aplicado de manera primaria. De acuerdo con Francis (2017) “los datos se reutilizan cuando se recopilan para un uso y luego se usan una segunda vez. Se reutilizan cuando el segundo uso tiene un objetivo diferente al primero”. Sin embargo, el objetivo primigenio del uso de los datos quedará sujeto a la propia naturaleza de la investigación que se desea desarrollar, pues el sentido lineal de la reutilización de los datos tiene que ver con el análisis intelectual que se realiza a los datos para su posterior reutilización. En este sentido, el uso de los datos queda determinado por su naturaleza temática y tienen una relación directa con la demanda informativa que manifiesta un determinado investigador al momento de responder las preguntas que se plantea dentro de su investigación.

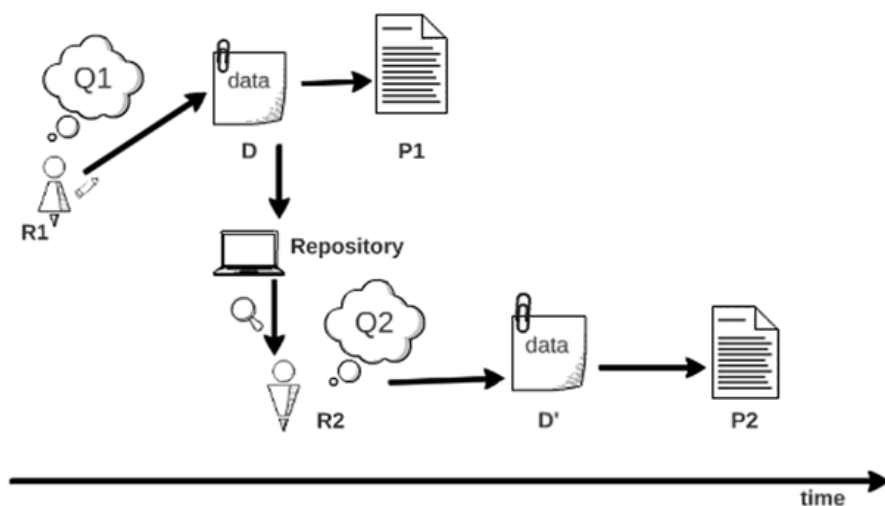


Fig. 4. Linear reuse model. Fuente: van de Sant, et al. 2019. <https://datascience.codata.org/articles/10.5334/dsj-2019-022/>

De acuerdo con el modelo lineal de reutilización (véase figura 4) propuesto por van de Sant et al., “se asume que los datos de investigación se crearon para un propósito específico, se analizaron y procesaron para cumplir con este propósito y finalmente se publicaron después de conducir a ciertos resultados de investigación. En este modelo, la reutilización solo ocurre después de que los datos se usaron para su propósito original (van de Sant, et al., 2019: p. 7)”.

De esta manera, el investigador 1 (R1) al plantarse una pregunta (Q1) comienza un proceso primario de investigación, por lo tanto, R1 recopila datos. El análisis de estos datos conduce a una publicación final (P1). Los datos se almacenan en un repositorio de datos de investigación y un nuevo investigador (R2) busca datos adecuados para abordar su pregunta de investigación (Q2). R2 puede transformar el conjunto de datos (D -> D') para adaptarlo a su pregunta y método, y publica los resultados en un artículo (P2) algún tiempo después de la primera publicación (P1).

Pues, en el modelo lineal de reutilización, intervienen factores como la temática de los datos, el perfil del usuario que va a utilizar a dichos datos, el propósito de la investigación que se va a desarrollar y el tiempo determinado que se tomará en cuenta para analizar a los datos y publicar los resultados de la investigación.

Por lo tanto, la reutilización de los datos de investigación forma parte de un proceso holístico que va más allá de la idea de publicar y visibilizar a los resultados de una determinada investigación. Se trata de dar apertura a los datos para que puedan ser contrastados, revisados, validados y reutilizados en un marco común de colaboración.

Los estándares para medir la calidad de las investigaciones científica y académicas cada día comienza a dar cuenta de la importancia de dar apertura a los datos de investigación, prueba de ello es la aparición de documentos basados en datos (como los data papers) que ponen de manifiesto la presencia de la reutilización en diversos elementos aplicados a los sistemas para acceder a los datos.

Sistematización de datos

En la actualidad, la sistematización de los datos (SD) es un proceso que va más allá de la creación de bases de datos y del uso de componentes computacionales para crear sistemas digitales con diversas características y funcionalidades. La SD es un concepto generalizador que toma en cuenta la interacción de los individuos con los datos que son registrados en perfiles de descripción (también conocidos como formularios), los cuales permiten almacenar a los datos como parte de un modelo.

En este sentido, un modelo de datos es “una notación para describir datos, que generalmente consta de tres partes: la estructura de los datos, las operaciones para obtener y actualizar datos dentro del modelo y las restricciones que, dentro del modelo, limitan los datos de varias maneras (Bresoud y Blanco, 2020)”. Las restricciones del modelo de datos permiten categorizar a los datos en un dominio particular. Este dominio puede ser temático, técnico y funcional, el cual toma como base la naturaleza de los datos y su contexto.

La SD ha sido foco de atención de múltiples investigaciones, sobre todo en campos como las ciencias de la información, la informática y las ciencias de la computación. Esto ha derivado en una serie de estudios que reflejan el corpus teórico y pragmático de la sistematización de los datos en múltiples entornos.

En el contexto digital actual, la SD toma notable relevancia debido al gran incremento de datos que se está aconteciendo. Los factores sociales, culturales y educativos han dado pauta para la generación de datos con múltiples características, algunas de ellas demasiado complejas para ser registradas y

representadas en una base de datos convencional. Prueba de ello, son los datos que se pueden desprender de contenidos digitales y de recursos audiovisuales.

De esta manera, los actuales modelos de datos buscan responder a las complejidades para registrar y representar a datos con características poco comunes o vistas con anterioridad. Esto permite vislumbrar una interrelación entre componentes conceptuales y técnicos entre diversas ópticas disciplinarias que son aplicadas para su efectivo manejo.

Además, existen diferentes tipos de dependencias de datos entre los atributos de una relación, como dependencias funcionales, transitivas y multivaluadas. Estas dependencias deben explorarse para desarrollar un modelo de datos que pueda usarse sin causar anomalías y problemas relacionados con el almacenamiento, la actualización y la eliminación de datos den un determinado sistema.

En el contexto de la investigación científica y académica, contar con modelados de datos flexibles y vinculables, permitirá establecer un mecanismo para la sistematización de los datos de investigación, fundamentado en principios interoperables que hagan posible la consolidación de un entrono virtual de investigación.

La SD no se trata única y exclusivamente de construir sistemas para almacenar y representar a los datos, sino de establecer un marco común para que estos datos puedan ser accesibles y reutilizados en el espacio digital y en un periodo de tiempo específico. Poder intercambiar datos, comunicarlos, extraerlos, reutilizarlos y revisarlos es esencial para la productividad científica, la colaboración y el descubrimiento mismo, por lo tanto, el modelado de datos de investigación es un aspecto que interactúa entre los diversos actores que harán uso de los datos de investigación como un recurso informativo con atributos propios. El análisis de estos atributos dará la pauta para la generación de un modelado de datos acorde a las necesidades de su sistematización.

De esta manera, aunque la naturaleza de los datos de investigación puede variar ampliamente según la disciplina, su importancia para la replicación, refutación o validación de los hallazgos u observaciones de un proyecto de investigación nunca ha estado en duda. Por este motivo, la sistematización de los datos debe permitir establecer un mecanismo para la consulta, acceso, descarga y reutilización de los datos, desde una visión holística y centrada en el usuario de dichos datos.

Oportunidades y desafíos de la reutilización y sistematización de los datos de investigación

En un contexto proliferado de datos e influenciado por el impacto de las tecnologías digitales, existen retos y oportunidades que se vislumbran al momento de contemplar la generación de un ecosistema para la reutilización y sistematización de los datos de investigación. Los cambios de paradigma en el contexto científico y académico, al momento de concebir la apertura de los datos de investigación, provocaran una serie de fenómenos, problemáticas y alcances que están relacionados con la democratización del acceso a los productos de investigación que fueron desarrollados en el sector público y patrocinados con fondos y recursos de una determinada sociedad. Pues, los datos de investigación, en la mayoría de los casos, son el resultado de la generación de investigaciones desarrolladas con fondos públicos.

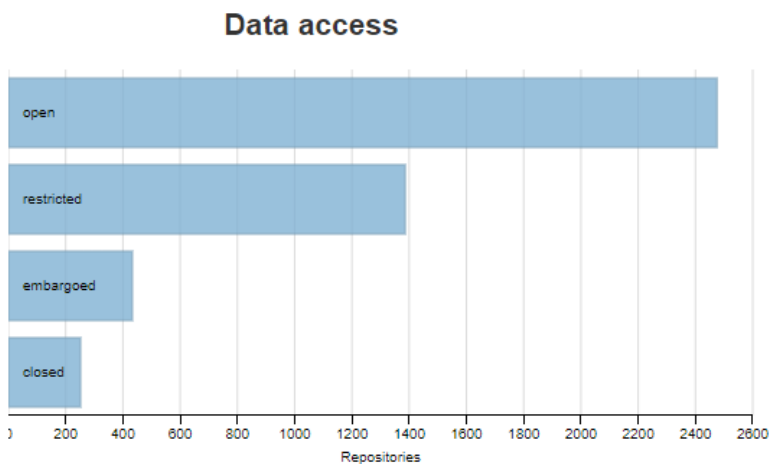


Fig. 5. Acceso a datos de investigación. Fuente: R3Data. Fecha de consulta: 03 de junio de 2022. Disponible en: <https://www.re3data.org/metrics/dataAccess>

El acceso a los datos de investigación es una variable que puede servir para identificar las fortalezas y debilidades a los que la reutilización de datos se enfrenta. En la figura 5, puede observarse el tipo de acceso que tienen los repositorios de datos de investigación registrados en R3Data. A la fecha; 2,475 repositorios presentan acceso abierto, 1,386 repositorios son de acceso restringido; 433 presentan acceso con periodo de embargo y finalmente 252 repositorios son de acceso cerrado. Estas estadísticas, permiten obtener una visión generalizadora de la postura que las instituciones académicas y de investigación están tomando, en el sentido de privilegiar el acceso abierto a los

datos de investigación, con lo cual se favorece la apertura de las investigaciones y el factor propio de su reutilización.

A pesar de la importancia del acceso a los datos para el progreso científico, la investigación sobre el impacto de limitar el acceso a los datos en la velocidad y dirección de la investigación científica es limitada. Aunado a ello, el acceso mejorado a los datos también democratiza la ciencia, al ayudar de manera equilibrada a los científicos del mundo en desarrollo y a las instituciones de menor rango a publicar utilizando datos de un determinado repositorio. Esta democratización, a su vez, aumenta la diversidad geográfica y temática de la investigación habilitada por el propio uso de los repositorios de datos. Por lo tanto, la accesibilidad a los datos de investigación pone de manifiesto la generación de políticas de datos que fomenten y fundamenten su reutilización. La variable del acceso a los datos, quedará influenciada por la capacidad de los investigadores para descargar y hacer uso de los datos en un determinado momento.

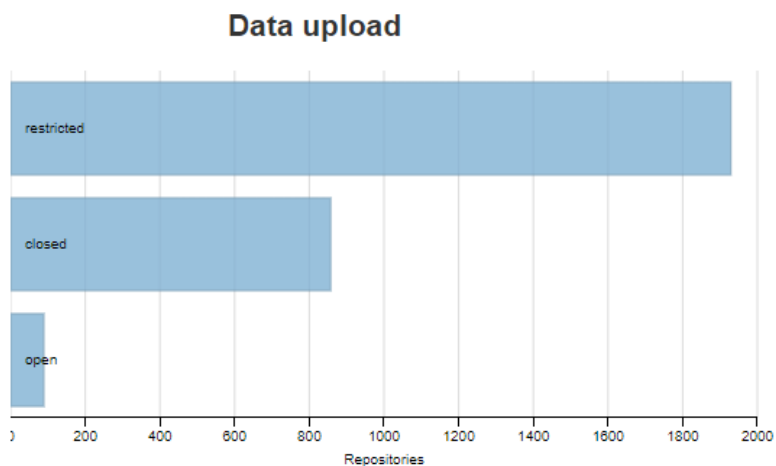


Fig. 6. Descarga de datos. Fuente: R3Data. Fecha de consulta: 03 de junio de 2022. Disponible en: <https://www.re3data.org/metrics/dataUploads>

De esta manera, en la figura 6, se aprecia que la descarga de los conjuntos de datos disponibles en R3Data presentan números desfavorables para fomentar una integral reutilización. Pues 1,929 repositorios restringen la descarga de sus conjuntos de datos; 857 repositorios emplean la modalidad cerrada de descarga y solo 89 repositorios ofrecen una descarga abierta.

El bajo número de repositorios de datos de investigación que hacen posible la libre descarga de los conjuntos de datos hace suponer que existe una falta

de desarrollo de políticas de gestión de datos de investigación que cubran los principios que debe gobernar a la reutilización de los datos. Pues sin una apertura al dato será imposible concebir su reuso. Plos One (<https://journals.plos.org/plosone/s/data-availability>) y Figshare (<https://figshare.com/>) en años recientes, han centrado su accionar en la generación de políticas consistentes que generen la accesibilidad y reutilización de los datos.

Por lo tanto, el reto principal de la reutilización de los datos de investigación consiste en convencer a las instituciones de investigación en participar en el modelo del acceso abierto, esto mediante su disposición para desarrollar políticas institucionales que fomenten la accesibilidad abierta a los datos de investigación.

Propiciar el desarrollo de un ecosistema de investigación fundamentado en los principios de apertura permitirá establecer intercambios de datos para estudiar los problemas urgentes que quejan a la sociedad. Por ejemplo, en tiempos de pandemia, “las redes de colaboración científica han transformado y modificado su accionar y la manera de desarrollar investigación al adaptar sus alcances en la aplicación de nuevas metodologías para el estudio del SARS-COV2 de manera integral pero también virtual (Ávila, 2022: p. 41)”.

Bajo esta premisa, desarrollar licencias específicas para los datos de investigación, será crucial para determinar el desarrollo de proyectos colaborativos basados en el procesamiento intensivo en datos. Establecer los alcances y limitaciones del uso de los datos en su faceta legal, económica y técnica permitirá sustentar los principios de la reutilización de los datos en el entorno digital.

Retos	Oportunidades
Cambio de paradigma en la manera de compartir conocimiento	Accesibilidad abierta a los datos
Generación de políticas centradas en la apertura de los datos	Visibilidad de las investigaciones desarrolladas con fondos públicos
Flexibilización de barreras económicas, legales y técnicas para el acceso a los datos	Reutilización de los datos de investigación
Definición clara de uso y reutilización de datos de investigación	Dar pie a nuevas investigaciones y descubrimientos

Valoración y fiabilidad de los datos de investigación	Elevar los estándares de la investigación, al permitir el escrutinio de los hallazgos y asegurar la integridad y reproducibilidad de la investigación
-------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------

Tabla 1. Retos y oportunidades de la sistematización y reutilización de los datos de investigación. Fuente: elaboración propia, 2022.

De esta manera, en la tabla 1 se exponen los principales retos y alcances de la sistematización y reutilización de los datos de investigación. Si bien las consideraciones expuestas en esta tabla no son exhaustivas, permiten estimar el comportamiento de los datos de investigación, sobre todo en condiciones actuales de crisis sanitaria, en donde los análisis y estudios se realizan con el propósito de obtener respuestas casi inmediatas para atender una problemática de dimensiones mayores, como es el caso de la pandemia que aún se sigue experimentando.

De esta manera, la sistematización y reutilización de los datos de investigación es un fenómeno con múltiples aristas. Desde aquellas que van del punto vista tecno-digital, hasta aquellas relacionadas con el factor social que caracteriza el uso de los datos para abordar fenómenos de amplias magnitudes. Por lo tanto, el uso de datos de investigación tiene un carácter multifacético en donde la figura de las instituciones de investigación es crucial para el desarrollo de políticas que sustenten el aprovechamiento y explotación de dichos datos.

Consideraciones finales

La reutilización de los datos de investigación es un tema con múltiples aristas. Desde aquellas que versan en la generación de políticas institucionales que fomenten dicha reutilización en el entorno académico y de investigación, hasta incursionar en el análisis del comportamiento de los datos al momento de utilizarse en un proceso formal de investigación.

La apertura a los datos de investigación abarca cuestiones que involucran la eliminación de barreras técnicas, económicas y legales para que puedan ser accesibles y reutilizados. Sin embargo, en la actualidad los principios que regulan estas condiciones de apertura se encuentran en desarrollo y en constante debate por parte de los actores que forman parte del proceso de publicación y divulgación del conocimiento científico.

Bajo esta premisa, las editoriales, los proveedores de datos y los investigadores deben trabajar en acuerdo mutuo para fomentar el desarrollo de un ecosistema que respete la propiedad intelectual de los creadores de los datos y fomente su reutilización de una manera armónica y apegada a los criterios éticos de la investigación.

La ética en el contexto de la investigación involucra la capacidad de reconocer y dar crédito a los creadores de una determinada obra intelectual. Los datos de investigación forman parte de este reconocimiento. A veces, los datos de investigación que involucran a personas no se pueden compartir públicamente debido al riesgo de violar la privacidad de los sujetos implicados en un determinado estudio. Sin embargo, incluso la información altamente confidencial puede compartirse de manera ética y legal si se proporciona más información sobre la protección de sujetos a través de los comités de ética y al momento de desarrollar la debida descripción de los datos de investigación.

En este sentido, la sistematización de los datos de investigación no es un asunto exclusivo de la tecnología computacional que se utilizará para dar accesibilidad a estos datos. Se trata de un principio que toma en cuenta la naturaleza de los datos para ofrecer una estructura descriptiva y representativa capaz de considerar el registro de los elementos intelectuales, informativos y creativos de dichos datos.

A su vez, la calidad de los datos es otro factor importante que influye en los juicios de confianza de los reutilizadores al momento de implementarlos en un proceso de investigación. Pues la calidad de los datos está positivamente relacionada con la confianza de los reutilizadores de los datos. La reutilización de conjuntos de datos existentes tiene mucho potencial, una vez que los investigadores adoptan los alcances de modelo para compartir y explotar a sus datos, tomando en cuenta los niveles de confiabilidad adecuados al momento de someter a sus datos a un proceso de citación.

Bajo esta premisa, los datos se reutilizan cuando se recopilan para un uso y luego se usan una segunda vez. Se reutilizan cuando el segundo uso tiene un objetivo diferente al primero. Si bien, la reutilización de los datos no es tema novedoso para el contexto académico y científico, en la actualidad debido al incremento en el uso de las tecnologías digitales para dar accesibilidad a los resultados de investigaciones desarrolladas con fondos públicos, es necesario establecer un marco común para asegurar que los datos tengan el reconocimiento de su creación y reutilización, mecanismos como la citación

de los datos es un ejemplo de la generación de marcos que están relacionados con el registro y descripción de los datos.

Finalmente, los esquemas de metadatos destinados a la descripción y representación de los datos de investigación son elementales para asegurar la interoperabilidad de los datos en el entorno digital. Pues a través de identificadores únicos y persistentes se puede asegurar la integridad de los datos y dar seguimiento a su reutilización en diversos campos disciplinarios y productos de investigación. De esta manera, se fomentará el desarrollo de un entorno virtual de investigación transparente y apegado a los principios de la universalidad.

Obras consultadas

- Ávila Barrientos, E. (2022). La investigación del SARS-COV2 mediante el uso de datos abiertos y grafos de conocimiento. En Georgina Araceli Torres Vargas (Coord.). *La pandemia por COVID-19: un acercamiento desde la bibliotecología y los estudios de la información*. (pp. 35-59). UNAM: Instituto de Investigaciones Bibliotecológicas y de la Información.
- Borgman, C. (2015). *Big data, little data, no data: Scholarship in the networked world*. The MIT Press.
- Bressoud, T. y White, D. (2020). Data Systems Models. En *Introduction to Data Systems*. Springer, Cham. https://doi-org.pbidi.unam.mx:2443/10.1007/978-3-030-54371-6_5
- Farnel, S. y Shiri, A. (2014). Metadata for Research Data: Current Practices and Trends. *International Conference on Dublin Core and Metadata Applications*. <https://dcpapers.dublincore.org/pubs/article/view/3714.html>
- Floridi, L. (2016). "Concepciones semánticas de la información". En Claudia E. Vanney, Ignacio Silva y Juan F. Franck (Eds.). *Diccionario Interdisciplinar Austral*, http://dia.austral.edu.ar/Concepciones_semánticas_de_la_información
- Francis, L. y Francis, J. (2017). Data Reuse and the Problem of Group Identity. En *Studies in Law, Politics, and Society*. (pp. 141–164). Emerald Publishing Limited. <https://doi.org/10.1108/S1059-433720170000073004>
- Lohr, S. (2015). *Data-ism: The Revolution Transforming Decision Making, Consumer Behavior, and Almost Everything Else*. Harper Collins.
- Pasquetto, I., Randles, B. y Borgman, C. (2017). On the Reuse of Scientific Data. *Data Science Journal*, 16 (8), 1-9. <http://doi.org/10.5334/dsj-2017-008>
- Van de Sandt, S., Dallmeier-Tiessen, S., Lavasa, A. y Petras, V., (2019). The Definition of Reuse. *Data Science Journal*, 18(1), 1-19. <http://doi.org/10.5334/dsj-2019-022>.