

Plan de Gestión de Datos en CONICET: análisis, experiencia y desafíos

GUILLERMINA ACTIS

LORENA CARLINO

Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)



Resumen

Principalmente se hace un recorrido de los avances que CONICET viene desarrollando en pos del cumplimiento de la Ley 26.899 y su reglamentación Res. MinCyT 753/16. Esta hoja de ruta se traza según las tendencias mundiales y estándares de interoperabilidad, las cuales enmarcan las decisiones que se han venido tomando en función de alinearse a la coyuntura de los organismos financiadores de proyectos de investigación y entidades académicas. Se muestran los análisis sobre mejores prácticas y casos de éxito en el desarrollo de herramientas para la elaboración de planes de gestión de datos en el mundo, y hacia dónde se dirige CONICET, incluyendo el giro hacia el *machine-actionable* de las herramientas más utilizadas en los últimos años. Se describe además la experiencia y sus resultados del PILOTO PGD del CONICET realizado en Agosto de 2016, dirigido a una convocatoria de financiamiento institucional, los Proyectos de Unidades Ejecutoras (P-UE). Finalmente, se han enumerado los retos y problemáticas que se presentan respecto a la necesidad de gestionar datos de investigación, dado que tienen un carácter relativamente reciente para las instituciones.

Palabras clave

Plan de gestión de datos (PGD); datos de investigación; gestión de datos.

Gestión de Datos Abiertos de Investigación (DAI)

Los datos de investigación son aquellas representaciones de entidades que quienes realizan actividades científicas y tecnológicas definen como evidencia, para el respaldo de sus hipótesis y resultados. La definición de lo que representa un dato en una investigación tiene por lo tanto un carácter contextual, y puede ser clasificado por su metodología de recolección o generación (observacional, experimental, computacional), por el estado en el que se encuentra de procesamiento (primario, procesado, derivado), por su origen (analógico, digital, digitalizado), entre otras. También pueden ser utilizados para la investigación datos que no provienen de o se generan en la misma, sino que se trata de registros (gubernamentales, institucionales, culturales) que los investigadores utilizan y procesan como evidencia (Borgman, 2015).

El carácter *abierto* de los datos de investigación, por su parte, deviene en la disponibilidad de los mismos para su uso, reutilización y redistribución, con la eventual y única condición de atribución para sus creadores y/o de compartir los resultados derivados bajo las mismas condiciones (LERU, 2014).

La gestión de datos no es una meta en sí misma, sino la serie de pasos y requisitos que permiten la reutilización de datos y la generación de nuevos conocimientos. Una buena administración de datos colectados en una investigación incluye la noción de "cuidado a largo plazo" tanto de *valiosos activos digitales* como de piezas físicas únicas, con el objetivo de que sean descubiertos y reutilizados en investigaciones posteriores. Los investigadores de todas las disciplinas siempre se han preocupado por sus datos, como base de la evidencia necesaria para la justificación de sus descubrimientos, sin embargo, en los últimos años la necesidad de gestionarlos se ha incrementado debido a dos fenómenos: el advenimiento de grandes cantidades de datos -*big data*-, con la digitalización de todas las disciplinas científicas por los avances tecnológicos; y el foco que han puesto las agencias de financiamiento y entidades gubernamentales en la necesidad de preservar los resultados de las inversiones públicas¹.

¹ La preocupación de las entidades gubernamentales por la forma en que los datos se vuelven disponibles y son compartidos se manifiesta desde hace varias décadas. En 1997 un reporte del National Research Council de Estados Unidos establecía que compartir datos de forma completa y era el fundamento de una ciencia "saludable", y un aspecto vital para garantizar el progreso nacional y maximizar los beneficios sociales a nivel global (National Research Council. 1997, p. 2).

En este contexto, se afirma que un buen manejo de datos facilitará y simplificará un proceso continuo de descubrimiento, evaluación, reutilización y evolución de la ciencia, bajo un paradigma de apertura que genere mayores beneficios por su carácter colaborativo². Para alcanzarlos, se ha puesto el foco especialmente en la posibilidad de que otros puedan no solo tener acceso sino también hacer uso efectivo de aquellos datos producidos en el contexto específico de un grupo de investigación y una disciplina particular.

Así, el trabajo de distintos grupos que promueven la apertura de datos de investigación se ha orientado a proporcionar cierta claridad en torno a los objetivos de una buena gestión de los mismos. Para ello en el año 2014 se realizó el evento *Jointly Designing a Data Fairport* en Leiden, Países Bajos, que reunió a investigadores, académicos del ámbito privado y público interesados en superar los obstáculos que existen en la localización y reutilización de datos de investigación. La reunión concluyó con la elaboración de un conjunto de principios fundamentales que establece que todos los datos de investigación deberán ser Localizables, Accesibles, Interoperables y Reutilizables (FAIR por sus siglas en inglés) tanto para máquinas como para personas (Nature, 2016).

La planificación de la Gestión de Datos

El primer aspecto de la gestión de los datos es la planificación de la misma. Si bien algunas disciplinas y campos científicos tienen larga trayectoria en el manejo de la evidencia que recolectan para sus investigaciones, el uso de nuevas tecnologías y los mandatos de las políticas institucionales han introducido la necesidad de considerar entre las actividades de investigación la previsión sobre las prácticas de generación o recolección y preservación de los datos. La experiencia ha demostrado que la dispersión de los datos aumenta el coste de procesamiento y la preservación, y que la inversión disminuye cuando se dispone de depósitos bien conformados (ACRL, 2014).

² Entre los beneficios que genera compartir datos de investigación, Borgman (2012) señala cuatro lógicas de motivación presentes en las argumentaciones que ofrecen tanto investigadores como financiadores y quienes elaboran políticas públicas de ciencia y tecnología: 1) reproducir o verificar investigaciones, 2) volver disponibles para el público general los resultados de investigaciones públicamente financiadas, 3) permitir la realización de nuevas preguntas, 4) avanzar el estado actual de la investigación y la innovación.

Si bien en muchos campos disciplinares compartir los datos es considerado una buena práctica científica y parte del ciclo de trabajo habitual, en muchos otros la preocupación por estos y su disponibilidad es reciente. Las instituciones gubernamentales reconocen la importancia de preservar y volver disponibles —cuando esto no implica riesgos— los resultados de las actividades públicamente financiadas.

En Argentina se estableció el acceso abierto a los resultados de las investigaciones, a través de la Ley 26.899. En línea con la Declaración de Berlín³ —la cual reconoce explícitamente a los datos como objeto del Movimiento de Acceso Abierto— la ley argentina establece que deben ser depositados en acceso abierto tanto las publicaciones como los *datos primarios* que sustentan las investigaciones públicamente financiadas, total o parcialmente.

El mandato de la ley también indica que las instituciones en las que las actividades se realizan deben garantizar, a través de repositorios digitales propios o compartidos, el acceso a estos resultados. Mediante la reglamentación de la ley en la Resolución 753/2016 se establece que las instituciones serán responsables “*sobre los procesos de administración, almacenamiento, conservación, preservación digital y supervisión de los datos primarios y de la producción científico-tecnológica resultante de los proyectos de investigación*”(Cap. I, Art. 1, ap. c).

Las instituciones que financien actividades de ciencia y tecnología deberán incluir en sus requisitos la exigencia de difundir en acceso abierto a través de los repositorios institucionales *los datos que pudieran generarse durante y como consecuencia de los proyectos de investigación financiados*(Cap. II, Art. 4.6). Como resultado de esto, los investigadores deberán por su parte explicitar en sus solicitudes de financiamiento los momentos del proyecto en los que esperan generar su producción científica (incluyendo los datos). Así, se entiende que es un requisito de la ley la planificación de los datos de investigación y su gestión, tanto para los investigadores como para las instituciones, que deben desarrollar un modelo de PGD a ser utilizado en los proyectos que estos inicien y durante el desarrollo del mismo (Cap. IV, Art. 16).

³ Disponible en <https://openaccess.mpg.de/Berlin-Declaration>.

Actualmente, la mayoría de las entidades financiadoras de proyectos de investigación exigen la presentación de un documento que evidencie una previsión del tratamiento de los datos colectados.

Un PGD es un documento que describe el tratamiento que van a recibir los datos de investigación recopilados o generados en el curso de un proyecto de investigación y después de que se haya terminado.

El PGD es en esencia una herramienta que facilita y acompaña la gestión de datos, describe el ciclo de vida de todo el conjunto de datos (dataset) que será recogido, procesado o generado por el proyecto de investigación.

El PGD no es un elemento fijo sino que evoluciona adquiriendo más precisión y entidad durante el período de vigencia del proyecto.

La comunidad científica mundial ha optado por el uso de aplicaciones informáticas que le faciliten la elaboración y mantenimiento del plan de gestión de datos.

Ventajas para la institución y para el investigador

Al considerar al Plan de Gestión de Datos como una herramienta de planificación se evidencian las siguientes ventajas tanto para el investigador como para la institución:

- **Ventajas para el investigador:**

Una correcta gestión de datos es fundamental para el éxito de la investigación, porque permite:

- Garantizar la integridad de la investigación, previendo la reproducibilidad de la misma y su verificación.
- Cumplir con los requisitos de agencias de financiación y las buenas prácticas.
- Aumentar la eficiencia y la calidad de la investigación, evitando la duplicación de tiempo y recursos a largo plazo.
- Mejorar la protección de datos y minimizar el riesgo de pérdida de los mismos, al tomar medidas y planificar acciones de antemano para su resguardo y anonimización, y depositándolos en entornos seguros con respaldo institucional.
- Garantizar la conservación y preservación de los datos de la investigación

- Evitar duplicaciones y labores innecesarias, al volver disponibles los recursos para su reutilización por otros.

- Ventajas para la institución:

La gestión desde el punto de vista de las instituciones donde se desarrollan actividades de investigación es clave porque no son solo espacios de creación de los datos sino que pueden volverse usuarias de los mismos, adquiriendo:

- Conocimiento previo de necesidades que surgirán en la gestión de datos a las que por ley se debe dar respuesta.

- Información más clara y estandarizada -machine readable, y con potencial aplicación de técnicas de minería de texto y datos- de las líneas de investigación y proyectos en desarrollo, para la asignación de recursos de forma más estratégica.

- Potencialidad de vinculación con el medio socio-productivo y/o actores gubernamentales, tanto al volver disponible la información que es base de desarrollos científicos y tecnológicos.

PGD en el mundo: estado actual y tendencia mundial

En los últimos años, varias organizaciones científicas en el mundo han emitido declaraciones y políticas que subrayan la necesidad de un inmediato archivo de los datos, y es así como algunos organismos de financiación han comenzado a exigir la presentación de un plan de gestión de datos (PGD), donde se establezcan además las medidas para volverlos abiertos. Estas declaraciones de las principales agencias de financiación de la investigación demuestran que la ética del intercambio de datos es esencial para maximizar el impacto y los beneficios de la investigación (ACRL, 2014).

En el contexto europeo, la principal entidad financiadora de investigación, la Comisión Europea, estableció un piloto en el año 2013 para su fondo Horizon2020, a través del cual un conjunto de proyectos prepararían PGD en el período 2014-2015 con las propuestas de investigación de sus solicitudes de financiamiento, y luego depositarían los mismos en repositorios de datos recomendados en el marco del proyecto OpenAire. La Comisión Europea creó un modelo guía de PGD, a través de la plataforma del Digital Curation Centre —DMPOnline—, que a la vez brinda apoyo en la elaboración de los planes y capacitaciones en la gestión y curaduría de los datos.

En los Estados Unidos también ha surgido en los últimos diez años iniciativas para la presentación de un PGD o como requerimientos obligatorios para el financiamiento de proyectos de investigación, como es el caso de la [NSF \(National Science Foundation\)](#).

Herramientas para la planificación de la gestión de datos

DMPTool

El [DMPTool](#) es una herramienta desarrollada en código abierto, creada a partir de la asociación de ocho instituciones⁴ estadounidenses a principios de 2011 como respuesta a los requisitos de las principales entidades financiadoras de actividades c&t⁵. Esta plataforma online permite la creación de PGD y provee de guías y recursos generales e institucionales para que el investigador pueda confeccionar su plan y, si lo desea, compartirlo (ya sea sólo con miembros de su institución o públicamente).

La herramienta también facilita el trabajo colaborativo, permitiendo que un investigador designe a un colaborador para trabajar sobre el plan, y permite la revisión por pares, a través de la funcionalidad de Revisión Formal e Informal de los PGD generados.

Además de su configuración para usuarios, permite que las instituciones elaboren sus modelos de PGD e incorporen las guías de los mismos, a los que sus investigadores tienen acceso a través de la identificación institucional.

DMPOnline

A partir de la estrategia del gobierno británico de financiar desde 2001 el impulso a la creación de infraestructuras digitales científicas (e-infrastructures & e-Science) se comenzaron a desarrollar proyectos en el

⁴ Las instituciones originales que colaboraron en la creación de la herramienta fueron: University of California Curation Center (UC3) de la California Digital Library, DataONE, Digital Curation Centre (UK), Smithsonian Institution, University of California, Los Angeles Library, University of California, San Diego Libraries, University of Illinois, Urbana-Champaign Library, University of Virginia Library

⁵ Los National Institutes of Health (NIH) añadieron el requisito de incorporar PGD a los financiamientos de más de U\$50.000 en el año 2003; la National Science Foundation (NSF) requería que los datos resultado de investigaciones que financiaba fueran compartidos desde el año 2001 pero el requisito se reforzó a partir de la evaluación por pares de los PGD como parte de las solicitudes de financiamiento (Borgman, 2012, 1060).

Reino Unido, entre ellos la creación de un Centro Nacional de e-Ciencia de un consorcio de universidades. El mismo, junto a otras dos instituciones⁶, ante la recomendación de Jisc⁷ de desarrollar un único centro nacional que se encargará de brindar servicios de curaduría digital, crearon el Digital Curation Centre (DCC).

DCC desarrolló una plataforma online desde la que se permite la elaboración de PGD ya sea con las guías provistas por este centro, o completando las plantillas con los requisitos de las instituciones educativas y financiadoras asociadas que solicitan a sus investigadores PGD en formatos específicos.

El futuro del DMP: DMP Roadmap/DMP Machine-actionable

El DCC y el equipo UC3 de la Biblioteca Digital de la Universidad de California han desarrollado las herramientas DMP Online y DMP Tool respectivamente para la planificación de la gestión de datos.

Como se dijo anteriormente, el DMPonline (DCC-UK) y el DMPTool (CDL-US) eran hasta hace poco un recurso para los investigadores que requerían la creación de planes de gestión de datos (DMPs). A partir de comienzos del año pasado, ambas instituciones se han unido para compartir experiencias⁸ y han decidido formalizar la asociación para desarrollar conjuntamente y mantener una única plataforma de código abierto para DMPs.

Pretenden así extender su alcance, mantener bajos los costos y avanzar hacia las mejores prácticas, generando así una participación en un ecosistema de ciencia abierta verdaderamente global.

⁶ Las Universidades de Edinburgh y Glasgow (en las que se alojaba el National e-Science Centre), UKOLN en la Universidad de Bath, y STFC, que tenía a cargo los Laboratorios Rutherford Appleton y Daresbury.

⁷ Joint Information Systems Committee, es una compañía sin fines de lucro abocada a brindar soporte y asesoría a instituciones de educación superior e investigación en recursos digitales, conexión y servicios tecnológicos.

⁸ La puesta en común de estas experiencias llevó a que los equipos de ambas herramientas acordar una serie de temas recomendables para incluir en un PGD: <https://github.com/DMPRoadmap/roadmap/wiki/Themes>. Para más información sobre los objetivos de la fusión de las herramientas y potencialidades para la ciencia abierta, ver Simms et. al. (2016). Disponible en http://riojournal.com/articles.php?id=8649&instance_id=3072183.

La nueva plataforma tiene como objetivo lanzar el DMP Roadmap en el segundo trimestre de 2017, implementando seguidamente y a corto plazo mejoras centradas en un DMP machine-actionable.

DMP machine-actionable

Los grupos de trabajo que se encuentran implementando la fusión de las dos herramientas tienen una visión particular sobre el futuro propósito de los PGD. Además de constituir una herramienta para la planificación del manejo, curaduría y preservación de los datos -tanto por parte de los creadores de estos como de las instituciones que tienen intereses en los mismos- se concibe a los PGD como potenciales fuentes de información para la generación de nuevos conocimientos.

En este sentido, la idea de que se conviertan en *machine-actionable* supone que el contenido de un PGD, su texto y los datos que contiene, puedan ser accesibles para la operación de los mismos a través de computadoras. A partir de esto, la visión del equipo de trabajo de la nueva herramienta prevé que los PGD dejen de ser simplemente un set de preguntas que los investigadores responden para obtener financiamiento, y en cambio se transformen en información estructurada de forma consistente, permitiendo su procesamiento a través de computadoras en función de ella. Esta propuesta permitiría que a través de la plataforma las instituciones puedan gestionar sus datos, las entidades financiadoras migren los PGDs que reciben por el otorgamiento de fondos, los proveedores de infraestructura puedan planificar el uso de sus recursos y los investigadores puedan descubrir más fácilmente la existencia de datos que les son útiles (CDL. 2016, p. 2).

La gestión de los datos abiertos de investigación en CONICET

Desde el CONICET se comenzó a trabajar sobre la gestión de datos en función del mencionado mandato de la Ley 26.899. A mediados del año 2014 se constituyó un equipo de trabajo configurando un Comité Asesor, compuesto por miembros de distintos sectores del Consejo, para empezar a trabajar en el desarrollo y en agosto de 2015 se lanzó el repositorio de publicaciones CONICET Digital, para el cual fue necesario establecer Políticas Institucionales de Acceso Abierto. En esta línea de acción, luego de la puesta en producción del repositorio institucional y en concordancia con la

Ley, se comenzó a avanzar sobre el segundo tipo de objeto de la misma: los datos primarios de investigación.

Para ello, se está trabajando en los lineamientos y aspectos que harían al cumplimiento del mandato y cuáles serían las responsabilidades tanto de los investigadores y personal del CONICET como de la institución.

En primer lugar, a principios del año 2016 se definió que sería necesario para el depósito de los datos primarios contar con un plan de gestión de los mismos. Siguiendo los modelos internacionales presentados, se trabajó en un conjunto de preguntas y requisitos a completar por los grupos de investigación en el momento de la elaboración de un proyecto. Luego, como un cuestionario tabulado, se envió a un *grupo piloto* durante la convocatoria a Proyectos de Unidades Ejecutoras de 2016.

A partir de esta experiencia piloto se avanzó en la definición de las necesidades técnicas que tiene el organismo y en las acciones para una correcta implementación del mandato de la Ley. De la mano de la Reglamentación de la misma en el mes de noviembre de 2016 se confirmó la dirección tomada hacia la obligatoriedad de presentación de PGD, y se comenzó a trabajar entre las áreas en la elaboración de Políticas Institucionales de DAI.

PGD en CONICET: acciones implementadas

En el año 2016 desde CONICET se iniciaron los trabajos para el análisis y definiciones de los aspectos de obligatoriedad de la apertura de los datos de investigación.

La primera etapa constituyó en la experiencia piloto dirigida a una convocatoria de financiamiento institucional, los Proyectos de Unidades Ejecutoras (P-UE). Estos proyectos debían elaborarse según consideraciones estratégicas de los institutos y centros de investigación y desarrollo del CONICET, para ser realizados en el plazo de cinco años con un financiamiento de hasta \$5.000.000⁹.

⁹ Estos proyectos tienen el propósito de promover investigaciones científicas y tecnológicas en distintas etapas: desde la etapa básica o de laboratorio, hasta la experimental e incluso de transferencia, promoviendo que las unidades ejecutoras integren para ello los esfuerzos de los distintos grupos e investigadores que las componen. Los proyectos deberían enfocarse en problemáticas locales, regionales o nacionales (identificadas en el Plan Argentina 2020 del MINCYT: <http://www.argentinainnovadora2020.mincyt.gob.ar/>). Las bases de la convocatoria actualmente en

De la convocatoria participaron 210 de las 217 unidades ejecutoras (UE) habilitadas para la presentación. De 275 ideas proyecto presentadas fueron seleccionadas 136 y finalmente 114 fueron los P-UE aprobados por el Directorio de CONICET.

En la instancia de las presentaciones de proyectos definitivos (las 136 ideas proyecto seleccionadas) se solicitó a un grupo piloto (y permitió el envío de forma voluntaria a quienes lo desearan) que completaran un cuestionario con una serie de opciones sobre la Gestión de sus Datos de Investigación en el marco del proyecto.

Se obtuvo respuestas de los titulares/directores pertenecientes a 40 proyectos, obteniendo así un panorama de cuáles eran los niveles de conocimiento y capacidades disponibles para las actividades de gestión de datos de investigación.

La información solicitada consistía en responder los siguientes aspectos:

- Qué tipos de datos generará y recopilará el proyecto.
- Qué estándares podrían utilizarse.
- Cómo serán compartidos los datos para su verificación y reutilización.
- Disponibilidad de los datos de investigación (cuándo).
- Cómo se conservarán y preservarán los datos.

En una sección final se incluyeron preguntas relacionadas con el Piloto, su contenido y conocimientos en general acerca de este tipo de herramientas. Esta información es actualmente muy relevante a la hora de encarar el proyecto definitivo del PGD.

Los resultados han demostrado que los investigadores tienen gran interés en el uso de herramientas que permitan la planificación del tratamiento de los datos y sus respuestas evidencian algunas necesidades que deben ser cubiertas a corto y mediano plazo:

- Desarrollar un repositorio de datos de investigación como entorno de localización y consulta pero también de *resguardo y preservación* de los datos de investigación.
- Contar con más infraestructura para proveer un espacio de almacenamiento para los datos de investigación.

desarrollo se encuentran disponibles en el siguiente enlace:
<http://convocatorias.conicet.gov.ar/proyectos-de-investigacion-de-ue-conicet/>.

- Dotar de equipamientos apropiados para mejorar la recopilación y gestión de los datos.
- Capacitar en tecnologías de la información, uso de dispositivos y alternativas para realizar resguardos, uso de herramientas y recursos para la investigación.
- Desarrollar material de consulta rápida (guías, manuales, glosarios) donde los investigadores puedan referirse para resolver cuestiones prácticas que hacen a la gestión de datos.
- Crear campañas de concientización sobre el acceso abierto y sus beneficios.
- Difundir la política de la gestión de datos.
- Crear un sitio web especial para el PGD.

PGD en CONICET: hacia dónde vamos

En el cumplimiento de los mandatos de la ley, se está trabajando en el desarrollo de herramientas para la gestión de datos de investigación y en la elaboración de su política institucional, el desarrollo de un repositorio de datos de investigación y en la integración de los sistemas, para lo cual se parametrizarán herramientas en código abierto y la adopción de las mejoras prácticas que se realizan en el mundo.

En el último año se han realizado distintos abordajes sobre la gestión de datos que, junto a la implementación del Piloto del PGD, y el análisis tanto de las experiencias internacionales que se han venido generando como de los mandatos que la institución tiene por ley, se han sumado a la necesidad de suministrar herramientas que permitan una buena gestión de datos. Para ello, se han determinado las siguientes pautas:

1. La obligatoriedad de presentación de un PGD regirá para proyectos aprobados a partir de 2017.
2. La primera versión del PGD deberá realizarse durante los primeros *meses* de marcha del proyecto (como un entregable).
3. En la planificación se deberán considerar la gestión de aquellos datos que sirvan para validar los resultados presentados en publicaciones científicas, y otros datos derivados de los proyectos que no sean publicados. Los datos que no deberán estar considerados para su apertura son:

- análisis preliminares, borradores de artículos científicos, planes para futuras investigaciones;
- secretos comerciales, información comercial, materiales confidenciales hasta el momento de su publicación, o información similar que esté protegida por la Ley;
- información médica o personal, así como información cuya divulgación constituya una clara invasión injustificada de la privacidad personal.

4. Además de la obligatoriedad de presentación al comienzo del proyecto, en fases sucesivas del mismo se deberán suministrar versiones más elaboradas y completas. El PGD deberá actualizarse por lo menos una vez hacia la mitad del transcurso del proyecto y otra al final para realizar los ajustes necesarios sobre los datos generados y los usos potenciales.

Los requisitos para completar el plan de gestión de datos serán provistos por CONICET en una Guía de Implementación para todas las áreas del conocimiento, pudiendo determinarse algunas salvedades para distintas disciplinas.

En este sentido, se ha realizado un análisis de factibilidad técnica para implementar el PGD CONICET a través de las herramientas más utilizadas y el DMP Roadmap. Si bien este aún no fue lanzado, su característica de código abierto permitió que se realicen análisis de sus avances de desarrollo a través del GitHub/Wiki, determinando que la herramienta contiene la mayoría de los requerimientos del CONICET¹⁰.

Conclusiones y desafíos

Los organismos de ciencia y tecnología, especialmente los de regiones y países en desarrollo, se encuentran frente a grandes oportunidades con la promoción de la Ciencia Abierta. La colaboración, a través de la apertura de

¹⁰ Actualmente, CONICET se encuentra en contacto con el equipo de desarrolladores de la nueva versión del DMP (formado por miembros del DCC-UK y de la UC3-USA). Se realiza a la vez un seguimiento frecuente de los avances del desarrollo, a la espera de su lanzamiento en el Q2 de este año.

los distintos aspectos de los procesos de investigación, permite obtener mayores beneficios de la inversión que realizan los Estados Nacionales.

Sin embargo, considerar a los resultados (tanto publicaciones como datos) de estas actividades como bienes públicos también representa grandes desafíos para estos organismos. Si bien existe gran diversidad entre los mismos, se observan aspectos clave en común, para los cuales es necesario consensuar soluciones, compartir capacidades, competencias y experiencias.

- **Infraestructura:** las instituciones deben comenzar a considerar entre sus responsabilidades el resguardo y preservación de aquellos resultados de las actividades en los que han invertido. Si bien la obligatoriedad de crear repositorios institucionales y la disponibilidad de fondos del SNRD ha generado avances en la creación de infraestructura, es necesario continuar en esta nueva etapa para asegurar que los datos sean almacenados de manera confiable y su origen y tratamiento esté registrado, de forma que los mismos puedan ser fácilmente reutilizados (ya sea con nuevos propósitos de investigación o desarrollo, o para la verificación de resultados). En este sentido, la infraestructura necesaria no es solo física sino también en términos de recursos humanos se debe trabajar por la definición de *nuevos perfiles e incorporación de profesionales de la información*, que puedan garantizar el correcto funcionamiento y aprovechamiento de las bases de datos y sus centros.

- **Capacitación para investigadores:** distintos estudios sobre datos abiertos de investigación y su potencialidad (Hey, Tansley, & Tolle, 2009; Tenopir et al., 2011; Sayogo & Pardo, 2013; Lämmerhirt, 2016; European Commission, 2015; European Commission, 2016; Treadway, Hahnel, Leonelli et. al., 2016) establecen entre los principales obstáculos la reticencia de algunos investigadores a compartir sus datos. Entre las principales razones de la misma se encuentran los temores a que los datos no sean adecuadamente utilizados o interpretados y que no sean reconocidos por su contribución en la creación de los mismos. Estos aspectos deben ser resueltos tanto para lograr que los datos de investigación se compartan como para que se utilicen. Es clave para ello que se capacite en cuestiones de citación, curaduría y documentación de información adicional para que los datos utilizados puedan ser atribuidos a sus creadores de forma adecuada, como para que su disponibilidad e integridad estén garantizadas lo antes posible.

- **Curatoría y control de calidad:** los investigadores y otros actores reutilizarán los datos siempre que estos estén disponibles de forma adecuada

y su calidad esté asegurada. Para ello será necesario que las instituciones establezcan mecanismos y protocolos mínimos de calidad en el resguardo y preservación.

- Falta de definiciones en estándares comunes: tal como establecen los FAIR Principles, los datos abiertos solo pueden ser utilizados si los mismos son accesibles y otros los pueden encontrar fácilmente. Para ello es necesario que tanto investigadores como instituciones aseguren que cierta información mínima se encuentre disponible en la web a través de sistemas interoperables. Si bien existen iniciativas de coordinación de estándares, el desafío es realizar un seguimiento de las mismas y promover las modificaciones que sean necesarias para asegurar la visibilidad de los recursos.

- Falta de recursos y personal de IT: la necesidad de gestionar datos de investigación tiene un carácter relativamente reciente para las instituciones. Es imprescindible que las mismas realicen un diagnóstico de las capacidades y recursos que deberán incorporar para hacer frente a los desafíos mencionados, y conviertan en prioritaria la incorporación de los mismos. En este sentido, también es un desafío para las instituciones abocadas a la formación de estos profesionales la actualización de perfiles y definición de nuevas curricula que permitan trabajar con entornos tecnológicos rápidamente cambiantes y volúmenes cada vez mayores de datos de investigación.

- Nuevos mecanismos de evaluación: si bien en el caso argentino por ley el depósito y su disposición en acceso abierto de los datos de investigación es obligatorio, es importante considerar que el cumplimiento representará una actividad nueva en muchos casos para los investigadores. En este sentido, el desarrollo de la misma debería, tal como la reglamentación de la ley sugiere, incorporarse a los circuitos de evaluación para que existan mayores incentivos a su cumplimiento. El desafío en esta dimensión radica en la incorporación de nuevos indicadores en las evaluaciones de promoción y desempeño, que supongan la consideración de los datos abiertos como otro resultado de las actividades de investigación.

Referencias bibliográficas

BORGMAN, C. L. (2015). *Big Data, Little Data, No Data: Scholarship in a Networked World*. MIT Press, Cambridge, MA.

- BORGMAN, C. L. (2016, October). Motivations for Sharing and Reusing Data: Complexities and Contradictions in the Use of a Digital Data Archive. Quello Lecture, Michigan State University. Retrieved from <<https://works.bepress.com/borgman/395/>>
- BORGMAN, C. L., GOLSHAN, M. S., SANDS, A. E., WALLIS, J. C., CUMMINGS, R. L., DARCH, P. T., & RANDLES, B. M. (2016). Data Management in the Long Tail: Science, Software, and Service. *International Journal of Digital Curation*, 11(1), 128-149. DOI: <<https://doi.org/10.2218/ijdc.v11i1.428>>
- CALIFORNIA DIGITAL LIBRARY (Octubre, 2016). "DMPRoadmap: Making Data Management Plans Actionable" - Grant Proposal for IMLS.
- DIGITAL CURATION CONFERENCE (2017). [Página web] <<http://www.dcc.ac.uk/resources/how-guides>>
- EUROPEAN COMMISSION. (2015). Public Consultation. "Science 2.0": Science in transition.
- EUROPEAN COMMISSION. (2016). European Open Science Cloud, 48(8), 2016. DOI: <<https://doi.org/10.1038/ng.3642>>
- EUROPEAN COMMISSION. (2017). Directorate-General for Research & Innovation. H2020 Programme. Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020. Version 3.2, 21 March 2017.
- HEY, T., TANSLEY, S., & TOLLE, K. (2009). The Fourth Paradigm: Data-Intensive Scientific Discovery. External Research, Microsoft Research, Redmond, USA. *Proceedings of the IEEE*, 99(8), August 2011.
- JONES, S. (2011). 'How to Develop a Data Management and Sharing Plan. DCC How-to Guides. Edinburgh: Digital Curation Centre.
- LÄMMERHIRT, D. (2016). PASTEUR4OA Briefing Paper: Disciplinary differences in opening research data.
- NATIONAL RESEARCH COUNCIL. (1997). Bits of Power - Issues in Global Access to Scientific Data. National Academy Press. DOI: <<http://doi.org/ISBN 0-309-05635-7>>
- PAVÓN, JUAN. (2014). Plan de Gestión de Datos en Horizonte 2020. UCM.
- SAYOGO, D. S., & PARDO, T. A. (2013). Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data. *Government Information Quarterly*, 30(SUPPL. 1), S19-S31. DOI: <<https://doi.org/10.1016/j.giq.2012.06.011>>
- TENOPIR, C., ALLARD, S., DOUGLASS, K., AYDINOGLU, A. U., WU, L., READ, E., ... FRAME, M. (2011). Data sharing by scientists: Practices and perceptions. *PLoS ONE*, 6(6), 1-21. DOI: <<https://doi.org/10.1371/journal.pone.0021101>>
- TREADWAY, JON; HAHNEL, MARK; LEONELLI, SABINA; PENNY, DAN; GROENEWEGEN, DAVID; MIYAIRI, NOBUKO; HAYASHI, KAZUHIRO; O'DONNELL,

DANIEL; SCIENCE, DIGITAL; HOOK, DANIEL (2016). The State of Open Data Report. Figshare. DOI: <<https://doi.org/10.6084/m9.figshare.4036398.v1>>

WILKINSON, MARK D., MICHEL DUMONTIER, IJSBRAND J. AALBERSBERG, GABRIELLE APPLETON, MYLES AXTON, ARIE BAAK, NIKLAS BLOMBERG, ET AL. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3:160018+, March 2016. DOI: <<http://dx.doi.org/10.1038/sdata.2016.18>>