



POLÍTICAS PARA LA CIENCIA ABIERTA Y LOS DATOS CIENTÍFICOS EN AMÉRICA LATINA

Documento de LA Referencia. Febrero de 2018.

LA Referencia es una plataforma de organismos gubernamentales de ciencia y tecnología de América Latina que impulsan el Acceso Abierto (AA) a la producción científica con énfasis en los resultados financiados total o parcialmente con fondos públicos. Se enmarca en las estrategias nacionales de Acceso Abierto, a través de los Repositorios Institucionales (RI), “ruta verde”, conformando así una Red Regional de Repositorios de Acceso Abierto a la Ciencia.

Esta red, trabaja en políticas de AA, tecnologías y directrices, tendientes a fortalecer los repositorios que ofrecen los resultados de la investigación en esta modalidad. Sus estrategias y líneas de trabajo, tienen como principio básico la convicción de que el conocimiento es un bien público y que gran parte de estos resultados han sido financiados con aportes públicos de las instituciones o países.¹

A nivel mundial, existen declaraciones de apoyo al AA desde el *Global Research Council*, el *G8 Science Ministers* y la Comisión Europea. A la vez, agencias de financiamiento públicas o semipúblicas han declarado explícitamente como condición de financiamiento que sus resultados deben ser publicados en acceso abierto. Por ejemplo, el programa Horizonte2020 de la Comunidad Europea requiere que los resultados y los datos científicos producto de su financiamiento estén disponibles a toda la ciudadanía.

En la actualidad, la tendencia desde los organismos de C y T de impulsar la Ciencia Abierta (CA), representa nuevos desafíos y metas para LA Referencia²

La ciencia abierta comúnmente se refiere a los esfuerzos para hacer que el resultado de la investigación financiada con fondos públicos sea más accesible en formato digital para la comunidad científica, el sector empresarial o la sociedad en general. La ciencia abierta es el encuentro entre la antigua tradición de apertura en la ciencia y las herramientas de las tecnologías de la información y la comunicación (TIC) que han reformado la empresa científica y requieren una mirada crítica de los responsables políticos que buscan promover la investigación a largo plazo y también la innovación.

¹ Por bien público se entiende la definición clásica que aplica al conocimiento y justifica gran parte de la I+D pública. Tradicionalmente son producidos por el Estado. Desde una mirada más económica, es un bien disponible para todos y el uso por una persona no merma el uso para los demás. En esencia, productos y servicios no-rivales y no-excluyentes con usualmente enormes externalidades positivas y dificultades de captura por un agente.

² Making Open Science a Reality. OECD, 15 Oct 2015 DOI: 10.1787/5jrs2f963zs1-en
<https://www.innovationpolicyplatform.org/content/open-science>

Esta definición general de CA proviene de la OECD y reconoce la tradición de la teoría de Bienes Públicos en este campo.²

Ciertamente este concepto debe ser entendido como un medio para mejorar la calidad e impacto de la investigación y no como un fin en sí mismo, tal y como indica el documento. Asimismo, la CA es un marco general de supuestos sobre cómo debiera ser la difusión del conocimiento y cubre áreas como ciencia ciudadana, datos científicos, revisión de pares y procedimientos abiertos, nuevas métricas, etc. Además, otros autores identifican escuelas o paradigmas para analizar el fenómeno, como la de infraestructura (que se ocupa de la arquitectura tecnológica), la de políticas públicas (que se ocupa del acceso a la creación de conocimiento), la de medición (impacto y mediciones alternativas), la democrática (que se refiere al acceso al conocimiento) y la pragmática (investigación colaborativa).³

Es evidente que los múltiples aspectos que conforman la CA deben ser construidos desde diversos frentes y actores. En ese contexto, dada la identidad y misión de La Referencia, se acuerda que su aporte a mediano plazo, sumado a sus esfuerzos existentes de Acceso Abierto a las publicaciones científicas, se ampliará a los datos científicos, a través de infraestructura, servicios, directrices, etc. Específicamente, impulsará acciones en un área de la CA: el del acceso abierto a los datos científicos, desde la mirada de la generación de bienes públicos (directrices metadatos, interoperabilidad, tecnología, políticas y apertura).

Convencidos de que el acceso abierto y gratuito a los datos de investigación, especialmente los financiados con fondos públicos, permitirá:

- Reproducir y validar los resultados de mejor forma Generar nuevas investigaciones y conocimiento a partir de los datos científicos abiertos ya existentes
- Dotar de mayor transparencia a los fondos aplicados a la investigación
- Preservar la integridad de la investigación
- Evitar la duplicación de esfuerzos

Asimismo, una característica esencial es que América Latina cuenta con legislaciones nacionales de AA (Perú y Argentina el 2013 y México el 2014), que incluyen a los datos científicos como parte de la política. Chile por su parte con

³ Fecher, B. and Friesike, S. (2013). Open Science: One Term, Five Schools of Thought. In: Bartling, S. and Friesike (Eds.), Opening Science New York, NY: Springer, pp. 17-47. Available at: http://book.openingscience.org/basics_background/open_science_one_term_five_schools_of_thought.html

CONICYT, está impulsando buenas prácticas de gestión de datos científicos y Brasil lanzó un Manifiesto para datos abiertos desde el Ibict. No obstante de ello, los sitios que relevan repositorios de datos científicos muestran una realidad muy incipiente en la región.⁴

Desde otro ángulo, en las últimas dos reuniones del Consejo Directivo de LA Referencia se determinó que deben realizarse acciones en el ámbito de la CA para lo cual ya se definieron un conjunto de decisiones que se incluyen en este documento (San Luis de Potosí, 2016; Buenos Aires, 2017).

Las siguientes páginas detallan los elementos de diseño que se deben considerar y las prioridades de acción. Es patente, además, que este no es un tema tecnológico sino un cambio cultural donde los datos científicos se reconocen como un resultado significativo del proceso de la investigación.

Modelo

La Región tiene particularidades respecto del resto del mundo para situar la estrategia de Datos Científicos:

- La cultura de la producción científica en A. Latina es no comercial. En general, su volumen en términos de Repositorios y revistas científicas es el “doble” en modalidad de AA de lo que es su participación en artículos publicados en bases de datos conocidas (Web of Science, Scopus).
- A diferencia del resto del mundo, el financiamiento de la I+D sigue proviniendo mayoritariamente del Estado: directo o indirecto.
- Una parte importante de los socios de LA Referencia controla varios aspectos de la cadena de valor. Desde el financiamiento de I+D, evaluación de la producción, apoyo a revistas científicas de excelencia, a la compra consorciada de los resultados producidos. Por ello, son posibles los diseños articulados de políticas públicas a nivel nacional.

⁴ R3DATA <http://service.re3data.org/browse/by-country/>
ICSU World Data System. <https://www.icsu-wds.org/services/data-portal>

Concepto

Dado el debate en curso, es imprescindible diferenciar “datos científicos primarios abiertos” de la tendencia denominada “datos abiertos gubernamentales”. Esta última apunta a colocar, en parte relevante, los datos de gobierno en forma accesible a todos los ciudadanos. El primero, además de su apertura, tiene complejidades especiales debido a la necesidad de estándares de metadatos, de privacidad, interoperabilidad, de identificadores persistentes, de reproducibilidad y preservación, entre otros.

Focalización

Es clave avanzar en definiciones operacionales que den foco a la acción regional. Para ello, LA Referencia ha decidido que, en una primera etapa, la prioridad de acción se concentrará en los datos científicos que sustentan y validan las publicaciones para profundizar la reproducibilidad de la ciencia, en los datos científicos que se recopilan o generan con fondos públicos de investigación y en aquellos que se basan en Planes de Manejo de Datos (DMP, Data Management Plan, por sus siglas en inglés) y que crecientemente formarán parte de las propuestas de investigación a ser financiadas con fondos públicos.

A la vez, en busca de soluciones, abordará temas relacionados con el software y herramientas necesarias para comprender y reutilizar los datos mencionados en el párrafo anterior, junto con estándares de metadatos interoperables ⁵. Otra área de intercambio de experiencias son aquellas cuestiones relacionadas con la armonización o definición de un set común de elementos que se solicitan a los investigadores en los PMD a fin de evitar duplicaciones de recursos a escala regional.

Finalmente, es clave cooperar, desde una mirada como la expresada en este capítulo, con las iniciativas tendientes a incentivar y proponer planes de gestión de datos en las instituciones que los producen.

⁵ Por ejemplo, el piloto de Horizonte2020 de la Comunidad Europea, señala : “The Open Research Data Pilot applies to both the data and any metadata, documentation, software and tools needed to understand and reuse the data. At a base level, the Open Research Data Pilot applies to all data needed to validate results in scientific publications, but researchers may also choose to share other curated and/or raw data that they feel has value to others and can be shared.”
<https://www.openaire.eu/public-documents?id=808&task=document.viewdoc>

Disciplinas y Metadatos

Si bien los datos científicos a nivel regional se producen en todas las áreas del conocimiento y ha sido un componente de toda política científica no privilegiar una disciplina sobre otra, existen en América Latina escasas iniciativas de datos científicos abiertos, con algunas excepciones (portales de datos; repositorios institucionales, mandatos y planes de gestión; legislaciones).

A la vez, observamos que existen redes internacionales consolidadas o con una tradición en este campo (por ejemplo: astronomía, biodiversidad, oceanografía), redes que integran muchos investigadores de la región, los que acostumbran depositar en sus plataformas los datos de investigación que generan. No obstante de ello, también se detecta por parte de la comunidad una resistencia a cambiar ese hábito de “depósito internacional” por el “depósito local” (ya sea en el repositorio institucional, como en portales temáticos nacionales) y, más fuertemente aún existe una resistencia a la apertura de los datos por parte de quienes no tienen esa costumbre, en muchos casos por temor al “quién y cómo se apropiará de los datos” con un sentido de propiedad sobre los mismos y en otros porque se ve como una carga más para la actividad del investigador.

En este sentido, si bien LA Referencia avanzará en el 2018 en su piloto de datos con un enfoque multidisciplinario, también seleccionará a iniciativas temáticas que ya estén funcionando en la región. Esto le permitirá trabajar en un escenario flexible que permita evolucionar con una mirada realista sobre la región. Incluso, quizás deba acercarse a los principales portales temáticos del mundo y trabajar juntos por una mayor articulación e interoperabilidad que facilite a las instituciones y países de la red, la obtención de su propia producción científica.

Finalmente, se considera que hay áreas donde sería importante una mayor articulación ya que tienen un fuerte impacto en el desarrollo regional y calidad de vida, como ser: Biodiversidad y Medio Ambiente.

Asimismo, aunque sea obvio, cabe destacar que la prioridad de LA Referencia es la articulación con los servicios y repositorios de datos científicos que son abiertos, interoperables, de carácter no comercial, y que brindan acceso abierto a los datos de investigación, aún con un posible embargo temporal.

Finalmente, en relación a los esquemas de metadatos, existen un conjunto sujeto a la especificidad de cada disciplina (ej. astronomía, genómica, biodiversidad, ciencias

sociales, etc....)⁶, LA Referencia ha decidido que, como recomendación general para la descripción de todo tipo de datos científicos a depositar en los repositorios, adoptará el esquema Datacite v. 4.0. Dicha decisión se sustenta en que este esquema es el más extendido mundialmente a la fecha y por ser el que permite una descripción suficiente para todo tipo de disciplinas, no obstante de ello anima a los repositorios temáticos a sumar, a este esquema básico y común, los elementos correspondientes a los esquemas más específicos que sean un estándar en su área disciplinar. Esta decisión, tiene como principal objetivo garantizar la interoperabilidad y evitar la duplicación de esfuerzos.

Revistas científicas

La acción en datos debe ser sinérgica y/o complementaria a las iniciativas de las revistas científicas. América Latina es diferente de otras regiones y se distingue por un modelo de AA que en su mayoría no tiene costos de APC. Por ello, un foco prioritario es diseñar alternativas que permitan en el ámbito tecnológico y de estándares de metadatos una interoperabilidad entre los repositorios de datos científicos y las publicaciones. De hecho, se vislumbra como necesario trabajar con los grandes portales de revistas de la región para facilitar los enlaces entre los “papers” y los datos y viceversa, explorando el diseño de alternativas que permitan en el ámbito tecnológico y de estándares de metadatos una interoperabilidad entre los repositorios de datos científicos y las publicaciones. En ese contexto vemos como muy valioso ampliar el uso de DOI para los datos y que puedan ser enlazados en las publicaciones.

También, la producción de la ciencia debe estar inserta en los procesos de investigación para evitar la duplicación de esfuerzos. Se deben avanzar y colaborar en propuestas donde “reportar” sea directo y sencillo para nuestros productores de investigación.

Licencias

El diagnóstico indica que en la región, aún existe un alto grado de desconocimiento sobre cuestiones relacionadas a Derechos de Autor y Licencias de Uso. Es por ello que se ha acordado, avanzar en algunas orientaciones o recomendaciones sobre el tipo de licencias más convenientes para los datos científicos para facilitar la reutilización con certidumbre jurídica.

⁶ <http://www.dcc.ac.uk/resources/metadata-standards>

En tal sentido, se recomienda el uso de licencias Creative Commons, que es el estándar de facto en América Latina y también el más extendido a nivel mundial. Asimismo, se sugiere que la alternativa más conveniente, para permitir la reutilización y a la vez garantizar el AA a largo plazo es Atribución 4.0 Internacional (CC BY 4.0) o Atribución-NoComercial-CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0)

Tecnología

El diseño de esta red, no es sencillo en cuanto a la localización de los nodos y los repositorios. En algunos casos, se ha optado por servicios centralizados en la “nube”, por ende, en cualquier ubicación en el mundo que responda a las necesidades y en otros casos, se ha optado por sistemas localizados en dependencias físicas o territoriales debido a legislaciones existentes, cautela respecto de la propiedad y privacidad de la información y cuestiones relacionadas con la soberanía de la información.

En tal contexto, es conveniente mantener el modelo flexible de LA Referencia que apunta a estructuras interoperables y federadas. De hecho, es aconsejable una aproximación basada en nodos nacionales de datos. Para ello es conveniente que la tecnología para la cosecha de metadatos de datos científicos en nodos nacionales sea de código abierto y transferible a los nodos que integran LA Referencia. Por ello, LA Referencia evaluará en el 2018 – 2019 el desarrollo de un cosechador de datos científicos, transferible para los nodos nacionales, que pueda cosechar validar y transformar metadatos desde diversos sistemas en desarrollo con protocolo estándar (OAI-PMH).

Otro aspecto es la tecnología de los repositorios de datos donde aún no hay tecnologías “dominantes” en la región para la implementación de estos repositorios como ha sido Dspace para los repositorios de publicaciones. Dependerá también en parte de las disciplinas y sus redes específicas. Pero, a fin de impulsar estas infraestructuras en las instituciones, consideramos que es conveniente que estas tecnologías sean de código abierto, gratuitas y transferibles. La experiencia en la región para publicaciones de repositorios ha demostrado que logran un mayor uso, innovación y apropiación. Finalmente evita que un cierre temprano de tecnologías y estándares pueda conducir a situaciones de “encierro” o “lock in”. Por ello, LA Referencia, continuará privilegiando la Interoperabilidad en el desarrollo de soluciones para datos científicos, trabajando con las plataformas existentes.

Incentivos

Dado que la CA es un cambio cultural, se considera necesario que las instituciones, organismos y los Estados implementen acciones, recomendaciones y mandatos que promuevan este cambio. Una acción que tendrá verdadero impacto y resultados en este sentido, es que las actividades relacionadas con la gestión y la difusión de los datos, científicos sean consideradas como un factor más en la evaluación de las carreras académica.

Por otra parte, es clave que las instituciones de investigación adopten recomendaciones explícitas o mandatos sobre la gestión de datos científicos. La experiencia muestra que en la medida que ello emerge, se crea o justifica también un impulso a las legislaciones. De hecho, volviendo sobre los modelos de evaluación de los investigadores, estos deben partir de las instituciones y transformarse en una aspiración común; de otra manera la acción pública no es el resultado de un diálogo constructivo, sino que se interpreta desde la imposición. El rol inicial para modificar los modelos de evaluación debe provenir de la comunidad universitaria y del ámbito científico.

Costos

Es una pregunta relevante quién debe financiar las infraestructuras y los costos recurrentes. En la medida en que se extienda el uso de los PMD probablemente surjan rubros específicos para ello en los proyectos, también aportes de determinados gobiernos en su territorio, pero resultará vital que las propias instituciones y organismos, aporten recursos para ello ; por ejemplo, en áreas de infraestructura como parte de la misión institucional de gestionar su producción.

No obstante de ello, existe un conjunto relevante de instituciones de CTI en la región que simplemente no cuentan con los recursos adecuados. Esta es una situación de borde y se requiere una fuerte colaboración y cooperación intra-regional e internacional, así como tecnologías factibles de implementar y transferir y recomendaciones basadas en mejores prácticas que eviten las duplicidades de esfuerzos. Esta condición, define una diferencia entre países desarrollados y América Latina. De hecho, es conveniente analizar la declaración de "Open Data in a Big Data World" de ICSU que introduce elementos y consideraciones en este ámbito de países en vías de desarrollo⁷.

⁷ ICSU (2015). "Open Data in a Big Data World": The International Council of Science (ICSU); the InterAcademy Partnership (IAP); the International Social Science Council (ISSC); and The World Academy of Sciences for the advancement of science in developing countries (TWAS). <http://www.icsu.org/news-centre/news/top-news/leading-science-groups-urge-global-accord-on-open-data-in-a-big-data-world>

En este sentido, LA Referencia, se propone como el instrumento que los países que la integran utilicen para avanzar más rápidamente, en conjunto y buscando soluciones con economía de escalas.

Globalización

En relación a la dimensión internacional, LA Referencia, desde enero del 2015, participa en coordinación con COAR (Confederation of Open Access Repositories) del proyecto OpenAIRE que es la plataforma de AAO para el programa Horizonte2020. Una de las tareas ha sido expandir la adaptación/adopción de directrices y servicios en América Latina para demostrar la factibilidad de interoperabilidad.

Hoy OpenAIRE también lidera parte de las iniciativas en ciencia abierta y datos científicos en Europa y se mantendrá este trabajo conjunto a nivel internacional que fortalece la inserción de nuestra investigación y ofrece beneficios y sinergias tangibles. A su vez, se deben explorar relaciones con otros referentes globales como ICSU-Codata, RDA, Datacite.

Esta integración de LA Referencia a redes globales y proyectos con otras regiones, tiene como resultado, además de un alto posicionamiento de la red, la activa participación en el diseño de estrategias, pilotos y toma de decisiones que tienen un impacto mundial.

Principios FAIR

Los elementos anteriores están alineados con la política de encontrable, accesible, interoperable y reutilizable “FAIR” por su sigla en inglés, que están siendo implementada en Horizonte 2020 y la Comunidad Europea, para el manejo de datos,⁸

“FAIR data management In general terms, your research data should be 'FAIR', that **is findable, accessible, interoperable and re-usable**. These principles precede implementation choices and do not necessarily suggest any specific technology, standard, or implementation solution.”⁹

⁸. Nature (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3, Article number: 160018 (2016) doi:10.1038/sdata.2016.18 <http://www.nature.com/articles/sdata201618>

⁹H2020 Program. Guidelines on FAIR data management.

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

De hecho, una reciente declaración de la Comunidad Europea sobre European Science Cloud, dice: “La implementación de los principios FAIR debe ser pragmática y neutral desde el punto de vista de la tecnología, abarcando las cuatro dimensiones: ubicación, accesibilidad, interoperabilidad y reutilización. Los principios FAIR no son estándares prácticos. Las áreas disciplinarias deben desarrollar sus nociones específicas de datos FAIR de forma coordinada y determinar el nivel deseado. Los principios de FAIR deben aplicarse no solo a los datos de investigación, sino también a los algoritmos, herramientas, flujos de trabajo, protocolos, servicios y otros tipos de investigación digital relacionado con los datos”¹⁰

Ejemplos de algunos componentes:

- Encontrable: uso de identificadores persistentes, uso de metadata enriquecida, registro en buscadores.
- Accesible: recuperación por identificadores usando protocolos estándares, abiertos, gratuitos e implementables.
- Interoperable: lenguaje de representación formal, accesible, compartido. Vocabularios que siguen principios FAIR.
- Reusable: licencias de uso claras y visibles; información sobre el origen de los datos; estándares utilizados en su recolección; autores; etc..

Conclusión

En base a lo anterior, se reafirma que la aproximación de LA Referencia se caracteriza en los siguientes elementos:

- Un modelo basado en la federación de nodos nacionales que apuntan a cosechar metadatos de datos científicos bajo acuerdos comunes de política, tecnología y directrices.
- Uso de estándares y tecnologías abiertas y transferibles en una aproximación no comercial.
- Generación de bienes públicos que facilitan la construcción de alternativas sin crear barreras de encierro tempranas.
- Conexión internacional con las iniciativas similares para avanzar en interoperabilidad y evitar la duplicación de esfuerzos.

¹⁰ EUROPEAN COMMISSION. DIRECTORATE-GENERAL FOR RESEARCH & INNOVATION. The Director-General.. EOSC Declaration. Brussels, 26 October 2017.
https://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf