

Conexión de los repositorios institucionales del CIMMYT al Repositorio Nacional de Ciencia y Tecnología de México

Connection of CIMMYT's institutional repositories to the Mexican National Science and Technology Repository

ARACELI ZÚÑIGA
ALEJANDRA TENORIO
JESÚS HERRERA
RICHARD FULSS

Resumen

El Centro Internacional de Mejoramiento de Maíz y Trigo llevó a cabo un proyecto de mejora de sus repositorios institucionales como acreedor de financiamiento de una subvención pública, otorgada por el Consejo Nacional de Ciencia y Tecnología de México y obtenida en proceso competitivo. En el desarrollo del proyecto se destacan dos innovaciones por una parte, su conexión con el Repositorio Nacional de Ciencia y Tecnología de México en simultaneidad, mediante un *middleware* que hace las veces de repositorio único a nivel lógico, haciendo transparente la existencia de dos repositorios físicos tras él. Por otra parte, la normalización de ORCID como identificador de autor persistente institucionalmente, junto con la adquisición y registro de más de dos mil de ellos.

Palabras clave: CIMMYT, repositorio, DSpace, Dataverse

Abstract

As a beneficiary of a public grant, awarded by the Mexican National Council of Science and Technology, and attained as a result of a competitive process by the International Maize and Wheat Improvement Center, this latter institution carried out an improvement project on its institutional repositories in which two innovations stand out. On the one hand, their simultaneous connection to the Mexican National Science and Technology Repository, thanks to a middle-

ware that operates as a unique repository at the logical level, making the two physical repositories behind it transparent. On the other hand, the normalization of ORCID as the institutional author persistent identifier, together with the compilation and recording of more than two thousands of them.

Key words: CIMMYT, repository, DSpace, Dataverse

1. Introducción

El Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT) ha habilitado históricamente mecanismos para preservar y dar difusión a su más valioso producto: la información creada por su personal científico, transmitida en diversidad de formatos. En 2011 se creó el primer repositorio digital del organismo, con diversidad de materiales bajo el sello institucional como artículos, actas de conferencias y libros de los cuales el CIMMYT es poseedor de los derechos correspondientes para poder licenciar su uso no comercial como bienes públicos globales.

Este repositorio, realizado con tecnología DSpace, se ha dedicado a publicaciones multimedio, mientras que para poner a disposición del público general tanto los conjuntos de datos como el *software* de investigación producidos en el centro se creó otro repositorio, esta vez usando tecnología Dataverse, en el año 2014. Ambos repositorios están siendo cosechados por el Repositorio Nacional de Ciencia y Tecnología de México (RN) desde el año 2018, gracias a la ejecución de un proyecto de mejora de la instalación institucional de repositorios, financiado en su mayor parte por el Consejo Nacional de Ciencia y Tecnología de México (Conacyt), como resultado de un proceso competitivo de adjudicación de recursos económicos

2. Desarrollo

2.1. Marco teórico

El fuerte incremento de documentos digitales de investigación en los últimos años, junto con la aparición de movimientos y políticas de Ciencia Abierta, ha propiciado la aparición y adopción de múltiples tecnologías de repositorio digital, como DSpace, CKAN,

Figshare, Zenodo, ePrints o EUDAT (Amorim *et al.*, 2017). Las entidades interesadas en ellos, además de considerar su instalación y puesta en marcha, han de considerar cuestiones como sus funcionalidades, características, sostenibilidad y compatibilidad (Hakopov, 2016). Es fundamental, además de la elección de las herramientas tecnológicas, la planificación para el desarrollo de los repositorios, teniendo en cuenta las pautas de gestión del conocimiento tanto generales como institucionales, la estructura de la información a resguardar dentro del repositorio, así como las materias o temáticas albergadas. Esto facilita que los usuarios puedan acceder a la información de manera rápida y de diferentes formas. Intrínseco a cualquier repositorio digital es que debe tener la característica de preservar la información (Hakopov, 2016).

En la actualidad, la preservación de la información y su difusión se consideran un importante proceso informativo, en donde las tecnologías de la información y las comunicaciones, así como los principios de acceso abierto (OA, *open access*), han sido fundamentales. Es de esta manera como los repositorios temáticos o institucionales han sido piedras angulares para el establecimiento del OA alrededor del mundo. Los repositorios digitales han devenido no solo en instrumentos para la preservación y gestión de la información (Borges, 2015), sino, además, en herramientas para la conservación de la memoria histórica; un ejemplo de lo anterior es la Biblioteca Virtual de Yucatán (2019), la cual alberga información de alrededor de 20.000 documentos. Con estos sistemas las instituciones han preservado y dado difusión a sus acervos, en especial las instituciones académicas y centros de investigación en diferentes países, como ejemplifican los casos de International Atomic Energy Agency (IAEA) (Hakopov, 2016) o el de Consultative Group on International Agricultural Research (CGIAR) (Yabowork, 2017).

De igual manera que en los ejemplos mencionados, el CIMMYT creó su primer repositorio institucional en 2011, con el sistema DSpace v1.7, donde se consideraría la iniciativa de acceso abierto, la cual se podría definir como acceso libre de forma gratuita, permanente y en línea, además de ser un medio que apoya a la visibilidad, para favorecer el acceso a la información (Piñeiro Molina, 2015). Con ello, CIMMYT daba acceso a la información de su acervo institucional, integrado por numerosos documentos propios de investigación y de memoria histórica.

El repositorio se conformó, en sus inicios, por cinco colecciones públicas, que se encontraban abiertas a todos los interesados por la información del CIMMYT, además de contar con cinco colecciones privadas, relacionadas con información exclusiva para el personal del centro, las colecciones y la información fue organizada conforme a los programas de investigación de la institución, en donde solo se resguardaban documentos de literatura científica. Sin embargo, con el tiempo el repositorio evolucionó, no solo en tecnología, alcanzando la versión 5.4 de DSpace, que actualmente es utilizado, sino también en organización y contenidos, eliminando todos los materiales que no fuesen de acceso abierto y añadiendo documentos multimedia como fotografías y vídeos.

Para 2014 se crearía el segundo repositorio institucional, utilizando el sistema Dataverse, versión 3.0, el cual se define como una aplicación web de código abierto, donde se conservan las premisas relacionadas con los sistemas de repositorios, como son: compartir, preservar y analizar los datos adquiridos de las investigaciones creadas por los científicos (Harvard University, Institute for Quantitative Social Science, 2020). En el caso específico de CIMMYT, para el resguardo de los conjuntos de datos generados dentro del centro.

Teniendo en cuenta que Dataverse se considera, más que un repositorio, una red de repositorios, el CIMMYT cuenta actualmente con siete repositorios dentro de su red, siendo algunos de acceso abierto e incorporando otros con algunas restricciones motivadas por los acuerdos con los donantes del centro con cuyos fondos se realizan determinadas investigaciones (Centro Internacional de Mejoramiento de Maíz y Trigo, 2020).

Para el CIMMYT, en su Plan Estratégico 2017-2022, dada la importancia que el CGIAR (consorcio del que CIMMYT es miembro) daría a los acervos digitales, se planteó lo siguiente:

El CIMMYT ha asumido el compromiso de administrar de manera eficaz y eficiente los Bienes Intelectuales para que los resultados de la investigación sean diseminados a gran escala y se produzcan mayores impactos. Apoyamos cabalmente los Principios de Administración de los Bienes Intelectuales y la Política de Acceso Abierto y Administración de Datos del CGIAR, que son esenciales para diseminar los bienes públicos internacionales, resguardar y pro-

mover el uso de los recursos genéticos y reforzar la capacidad de investigación.

Utilizamos la Administración de Bienes Intelectuales como herramienta de diseminación de los resultados de la investigación como bienes públicos internacionales mediante diferentes esquemas de licenciamiento, repositorios de acceso abierto, canales de información específicos, capacitación, alianzas público/privadas e investigación participativa. Alentamos y trabajamos con nuestros colaboradores para que implementen los principios de acceso abierto y datos abiertos. (Centro Internacional de Mejoramiento de Maíz y Trigo, 2017, p. 28)

La estandarización de los repositorios en el CIMMYT ha crecido paulatinamente, incorporándose diferentes elementos a lo largo del tiempo, como el lenguaje controlado de la Organización de las Naciones Unidas para la Alimentación y Agricultura (FAO), AGROVOC, que es un tesoro multilingüe, centrado en la alimentación, la nutrición, la agricultura, la pesca, las ciencias forestales y el medio ambiente (Agricultural Information Management Standards –AIMS–, 2018). Este cuenta con alrededor de 8 millones de referencias bibliográficas referentes a las ciencias agronómicas (Celli y Keizer, 2016).

Por otra parte, en México se publicaría la Convocatoria 2016 para Desarrollar los Repositorios Institucionales de Acceso Abierto a la Información Científica, Tecnológica y de Innovación, al amparo de la legislación nacional relacionada con la ciencia y la tecnología, además de la educación, cuyos objetivos eran el desarrollo, la vinculación y la diseminación de los productos científicos, tecnológicos y de innovación en territorio mexicano, donde el principal soporte serían las plataformas de acceso abierto. El Consejo Nacional de Ciencia y Tecnología asumió esta tarea, teniendo como uno de sus resultados la creación del RN, el cual busca permitir el acceso y la disponibilidad de información en texto completo en formatos digitales para la consulta de la población en general (Tena Espinoza, 2018).

El CIMMYT fue una de las instituciones dotadas con fondos en la mencionada convocatoria (Conacyt, 2017), desarrollando un proyecto de mejora de acuerdo a sus lineamientos. Entre las actividades ejecutadas, no solo se estableció la conexión de dos repositorios institucionales en simultaneidad al RN, por primera

vez en México, sino que también se inició la utilización sistemática de ORCID, ya que funge como un identificador que ayuda a distinguir a los investigadores, además de garantizar el crédito por sus investigaciones (Taylor & Francis, 2020); esto fue realizado de forma simultánea con un análisis sobre posibilidades de conectividad entre ORCID y DSpace, buscando la estandarización de autores (Fernández, 2016). Aun no siendo las únicas actividades desarrolladas, estas constituyen las principales innovaciones del proyecto, que se revisarán a continuación.

2.2. Descripción de la innovación

La principal innovación conseguida fue la conexión con el RN de los dos repositorios físicos del CIMMYT como si fuera un único repositorio lógico. Este proyecto fue el primero en México en llevar a cabo la conexión simultánea de más de un repositorio institucional al RN para su cosecha. Como innovación de carácter secundario y ámbito local, se estableció ORCID como el identificador de autor oficial de CIMMYT para sus repositorios. Previamente a este proyecto se disponía de solo unos pocos ORCID de científicos institucionales registrados y muchos de ellos no disponían de uno.

2.3. Proceso de implementación de la innovación

Para conseguir la innovación principal se produjo un *middleware* que, simultáneamente, unificaba la conexión a ambos repositorios y adaptaba sus metadatos al formato requerido por el RN sin modificarlos en origen, lo que permite conservar las políticas de catalogación del CIMMYT, aunque no sean totalmente coincidentes con las del RN (por ejemplo, en la manera de registrar los nombres de autor). De esta manera, el *middleware* se conecta a los repositorios de publicaciones y datos mediante sus interfaces API-SOLR, extrae los metadatos requeridos, los transforma al formato apropiado para su consumo por el RN, compone el XML en formato OpenAIRE con los metadatos transformados y lo expone a través de una conexión OAI-PMH que es explotada por el RN para la cosecha.

El plan inicialmente proyectado, ejecutado en la primera implementación del *middleware*, preveía la exposición de los meta-

datos de ambos repositorios de manera combinada a través de una única conexión OAI-PMH. Sin embargo, en las pruebas de cosecha el Conacyt informó que el RN todavía no estaba preparado para diferenciar elementos de literatura o datos que vinieran combinados en un único XML Open-Aire. Esto motivó una modificación al proyecto en tiempo de ejecución, de manera que el *middleware* pasó a exponer por separado los metadatos de los repositorios de publicaciones y datos, ello a través de dos conexiones OAI-PMH.

La innovación secundaria se consiguió gracias a la promoción del identificador ORCID entre los científicos del centro, colaborando con la organización a cargo de este identificador, que se prestó a enviar representantes al CIMMYT para llevar a cabo presentaciones y registros de nuevos autores. Por otra parte, el uso de ORCID favorece la visibilidad y el impacto de las publicaciones, debido a su amplia aceptación mundial, que lo ha llevado a ser un elemento vinculante entre producción científica y autores. Durante 2018, parte de los investigadores del CIMMYT crearon su identificador ORCID, por ser uno de los requisitos solicitados por parte de las editoriales en las cuales publicaban. Esto ocurrió en paralelo a las labores de promoción interna en las que se les dieron a conocer los beneficios de dicho identificador.

Asimismo, también se llevaron a cabo tareas de obtención de identificadores ORCID cruzando información existente en diversos sistemas y publicaciones sobre artículos publicados que lo incluyesen, junto con la corroboración de las afiliaciones de los investigadores para asegurar la correcta asignación de identificadores para los autores.

2.4. Evaluación de resultados

En el caso de la innovación principal, el resultado de las acciones ejecutadas fue la conexión de ambos repositorios institucionales al RN. Una vez superadas las pruebas de cosecha, el CIMMYT emitió al Conacyt un oficio explicitando que sus repositorios institucionales cumplen con los lineamientos técnicos y jurídicos requeridos, junto con los puntos de acceso OAI-PMH para su cosecha ininterrumpida. Este oficio se puede considerar la prueba del éxito de la innovación aplicada. Por otra parte, el hecho de que el RN ha estado cosechando de manera continua

cifras incrementales de elementos de ambos repositorios, como se ejemplifica en la tabla 1 con los datos correspondientes al segundo semestre de 2019, da fe práctica de la funcionalidad desarrollada.

Tabla 1. Cosechas por el RN de los repositorios del CIMMYT, segundo semestre de 2019.

	Dataverse		DSpace		Total	
	Nuevas cosechas	Cosechados totales	Nuevas cosechas	Cosechados totales	Nuevas cosechas	Cosechados totales
Julio	N/A	148	N/A	336	N/A	484
Agosto	32	180	65	401	97	581
Septiembre	1	181	21	422	22	603
Octubre	5	186	17	439	22	625
Noviembre	2	188	1	440	3	628
Diciembre	77	265	673	1113	750	1378

Fuente: elaboración propia.

Otro indicativo del éxito de la innovación incorporada es el reconocimiento por parte de la comunidad, constatado por la invitación hecha por el Conacyt a CIMMYT para que presentase el proyecto desarrollado en el seminario *Entre Pares*, en su edición de 2018 (Herrera de la Cruz, 2018). En cuanto a la innovación secundaria, en DSpace y a comienzos de 2018, solo se encontraban dos identificadores ORCID integrados en el RN. Para inicios de 2019, tras realizar la primera promoción del identificador y tener alrededor de 169 ORCID ingresados de manera manual, se consideró hacer una segunda promoción, esta vez implementando el programa Collect & Connect de ORCID (2020). Mediante este proceso y de la mano del envío de correos electrónicos al personal de la institución, se confirmaron 123 identificadores más.

Por otra parte, se integraron de manera manual los ORCID de investigadores y colaboradores no pertenecientes a CIMMYT, los cuales fueron obtenidos mediante el cruce de información entre las diversas publicaciones y las bases de datos. Para finales de abril de 2019 se conseguirían 2096 identificadores. La evolución

numérica mencionada certifica claramente el éxito de esta innovación.

3. Conclusiones

Como resultado de la ejecución de un proyecto de mejoras de sus repositorios institucionales, y como acreedor de financiación pública obtenida en proceso competitivo, el CIMMYT pudo llevar a cabo dos notorias innovaciones entre la diversidad de acciones planeadas en él. La principal innovación, consistente en la conexión en simultaneidad de los dos repositorios del centro al RN mediante un *middleware* que actúa como repositorio lógico único, tuvo como resultado la conexión efectiva, cuyo éxito se comprueba con la continua cosecha de nuevos elementos relativos tanto a publicaciones como a conjuntos de datos, por parte del RN.

La innovación de carácter secundario aquí tratada, logró su objetivo de instaurar ORCID como el identificador de autor persistente institucional, llegando a conseguir y registrar más de dos mil de ellos hasta la fecha, lo que permite aprovechar los beneficios intrínsecos de un identificador con aceptación mundial, toda vez que ayuda a evitar problemas de otros identificadores aceptados por el Conacyt, en su mayor parte locales al entorno de la Administración Pública mexicana pero desconocidos por entidades fuera del país. Ambas innovaciones pueden servir de modelo para otras instituciones, de cualquier país, que dispongan de varios repositorios cuya cosecha conjunta sea un objetivo deseable.

4. Referencias

- Agricultural Information Management Standards (AIMS) (2018). *AGROVOC tesaurus multilingüe de agricultura*. <http://aims.fao.org/es/agrovoc>.
- Amorim, R. C., Castro, J. A., Rocha da Silva, J. y Ribeiro, C. (2017). A comparison of research data management platforms: Architecture, flexible metadata and interoperability. *Universal Access in the Information Society*, 16(4), 851-862. <https://doi.org/10.1007/s10209-016-0475-y>.

- Biblioteca Virtual de Yucatán (2019). Mérida, Yucatán: Secretaría de la Cultura y las Artes. <http://www.bibliotecavirtualdeyucatan.com.mx>.
- Borges, L. d. C. (2015). *Repositório educacional emsaúde: homenagem à Virgínia Torres Schall*. https://www.arca.fiocruz.br/bitstream/iciict/23387/2/leandro_borges_iciict_tcc_2015.pdf.
- Celli, F. y Keizer, J. (2016). *Enabling Multilingual Search through Controlled Vocabularies: the AGRIS Approach*. En: Garoufallou, E., Subirats Coll, I., Stellato, A., Greenberg, J., Celli, F., Keizer, J. y Publishing, S. D. (ed.). *Metadata and Semantics Research* (pp. 237-248). https://doi.org/10.1007/978-3-319-49157-8_21.
- Centro Internacional de Mejoramiento de Maíz y Trigo (2017). *Plan estratégico 2017-2022: mejorar los medios de vida mediante ciencia aplicada al maíz y al trigo*. <http://hdl.handle.net/10883/19073>.
- Centro Internacional de Mejoramiento de Maíz y Trigo (2020). *CIMMYT Research Data & Software Repository Network*. <https://data.cimmyt.org/dataverse/root?q=&types=dataverses&sort=dateSort&order=desc&page=1>.
- Conacyt (2016). *Lineamientos Específicos para Repositorios*. https://osf.io/pr2yc/?view_only=e18df09edfcb4983a9f3b17ea8264a3e.
- Conacyt (2017). *Convocatoria 2016 para Desarrollar Repositorios Institucionales de Acceso*. <https://www.conacyt.gob.mx/index.php/el-conacyt/convocatorias-y-resultados-conacyt/convocatorias-direccion-adjunta-de-planeacion-y-evaluacion/convocatoria-2016-repositorios-institucionales-aaicti/14344-resultados-de-la-convocatoria-2016/file>.
- Fernández, C. (2016). *El soporte de ORCID en DSpace 5 (y superiores)*. Hablando de DSpace: [página web]. <https://www.arvo.es/dspace/tag/authority-control>.
- Hakopov, Z. (2016). *Digital Repository as Instrument for Knowledge Management*. Preprint. <http://eprints.rclis.org/29046>.
- Harvard University, Institute for Quantitative Social Science (2020). *The Dataverse Network*. <https://ddialliance.org/project/the-dataverse-network>.
- Herrera de la Cruz, J. (2018). *Llevando CIMMYT al Repositorio Nacional [diapositivas]*.
- León Betanzos, G. A. (2016). *Manual para administración y uso de DSpace*. Ciudad Universitaria: DGTIC-UNAM. <http://www.ru.tic.unam.mx:8080/bitstream/DGTIC/81872/1/Manual%20DSpace%20r.pdf>.
- Ley de Ciencia y Tecnología, de la Ley General de Educación y de la Ley Orgánica del Consejo Nacional de Ciencia y Tecnología (2014).

- https://www.dof.gob.mx/nota_detalle.php?codigo=5345503&fecha=20/05/2014&print=true.
- Lorenzo-Gil, E. y Braña-Ferreiro, E. (2012). *Administradores de Colección y Comunidad*. Hablando de DSpace: <https://www.arvo.es/dspace/administradores-de-coleccion-y-comunidad>.
- OpenAIRE (2015). *OpenAIRE Guidelines for Literature Repositories v3*. Open AIRE Guidelines. https://guidelines.openaire.eu/en/latest/literature/index_guidelines-lit_v3.html.
- ORCID (2020). *Collect & Connect for Publishing Organizations*. Member Support Center: <https://members.orcid.org/cc-publishers>.
- Piñero Molina, M. (2015). Los repositorios de acceso abierto como alternativa. *Revista Cubana de Información en Ciencias de la Salud*, 26(4), 330-346. <https://www.medigraphic.com/pdfs/acimed/aci-2015/aci154c.pdf>.
- Roman, L. R. (s. f.). *Manual de Uso DSpace IAPH*. (I. A. Histórico, ed.) https://repositorio.iaph.es/ayuda/Manual_DSpace_IAPH_Ayuda_consulta.pdf.
- Taylor & Francis (2020). *ORCID: how to include it in your online submission (and why you should)* (Taylor & Francis, ed.). <https://authorservices.taylorandfrancis.com/orcid-how-to-include-it-in-your-online-submission-and-why-you-should>.
- Tena Espinoza, M. A. (2018). *Desarrollo de repositorios institucionales como estrategia nacional para el acceso abierto. El caso de México* (pp. 85-104). Universidad de Salamanca. <https://dialnet.unirioja.es/servlet/articulo?codigo=6746330>.
- Yabowork, A. O. (2017). *Making CGIAR outputs open and accessible: the CGSpace collaboration*. *Knowledge Management for Development Journal*, 13(2), 23-33. <https://hdl.handle.net/10568/82764>.

5. Reconocimientos

El proyecto parcialmente descrito en este artículo fue financiado casi en su totalidad por el fideicomiso 2250-6, denominado «Fondo Institucional del Conacyt», bajo convenio de asignación de recursos con identificador FON.INST./296/2017. La parte no cubierta por el fideicomiso fue ejecutada con cargo a los presupuestos ordinarios del Centro Internacional de Mejoramiento de Maíz y Trigo.