



BIREDIAL-ISTEC
CONFERENCIA INTERNACIONAL SOBRE BIBLIOTECAS Y REPOSITORIOS DIGITALES DE AMÉRICA LATINA



X Conferencia Internacional de Bibliotecas y Repositorios Digitales

Modalidad virtual
25 al 29 de octubre de 2021



Instituto Nacional de Astrofísica (INAOE)
México



Red Mexicana de Repositorios Institucionales

BIRE **EDIAL**
ISTEC



Actas de la X Conferencia Internacional de Bibliotecas y Repositorios Digitales

Modalidad virtual

25 al 29 de octubre de 2021

Instituto Nacional de Astrofísica (INAOE)

México



Red Mexicana
de Repositorios
Institucionales

Actas de la X Conferencia Internacional de Bibliotecas y Repositorios Digitales :
BIREDIAL-ISTEC / Pablo César de Albuquerque ... [et al.] ; compilación de
Marisa Raquel De Giusti ; coordinación general de Marisa Raquel De Giusti. -
1a ed. - La Plata : Universidad Nacional de La Plata, 2022.
Libro digital, PDF

Archivo Digital: descarga y online
ISBN 978-950-34-2109-3

1. Bibliotecas Virtuales. I. Albuquerque, Pablo César de. II. De Giusti, Marisa Raquel, comp.
CDD 020.7

Coordinación general de la edición: Dra. Marisa R. De Giusti

Edición, maquetación y corrección de estilo: Esp. Analía V. Pinto

Diseño de tapa: DCV Lucas E. Folegotto

Revisión técnica y traducción del portugués: Prof. Carlos J. Nusch

Esta obra se edita bajo una licencia [CC-BY](#).



Índice

PROGRAMA	7
PRESENTACIÓN	19
CONFERENCIAS MAGISTRALES (RESÚMENES)	22
Estado de situación y avances en ciencia abierta y acceso abierto en Argentina Legislación, implementación, resultados, retos	23
¿Cómo el Acceso Abierto diamante puede contribuir hacia una evaluación responsable e inclusiva de la ciencia?	26
Datos Abiertos: Gestión de datos de investigación	28
Repositorios, plataformas abiertas y comunitarias: para la sostenibilidad, innovación y equidad en la comunicación científica y académica	31
MESAS DE DEBATE (TRANSCRIPCIONES)	33
Mesa 1: Políticas nacionales de Acceso Abierto en América Latina	34
Mesa 2: Evaluación y métricas	42
Mesa 3: Datos de investigación en abierto	58
Mesa 4: Infraestructuras tecnológicas para gestión de datos de investigación	69
Mesa 5: Experiencias de nodos nacionales de LA Referencia	84
PONENCIAS	104
EJE 1: COMUNICACIÓN ACADÉMICA, CIENTÍFICA Y CULTURAL EN ABIERTO	105
Modelo de gestión de ciencia abierta para la Universidad Autónoma de Zacatecas	106
Hacia la formación de investigadores en prácticas abiertas	106
Lineamientos institucionales sobre Ciencia Abierta: caso de la Universidad del Rosario en Colombia	125
Capacitação dos Profissionais de informação visando a gestão no Arca – Repositório Institucional da Fundação Oswaldo Cruz	141

Mapeamento dos projetos brasileiros em Ciência Cidadã e Biodiversidade	150
EJE 2: DATOS ABIERTOS	171
Colecciones públicas abiertas: propuesta de depósito digital de metadatos bibliográficos de las publicaciones oficiales de Bogotá	172
EJE 3: EVALUACIÓN Y MÉTRICAS ALTERNATIVAS	196
Datos bibliométricos para las Ciencias Sociales y las Humanidades: un método para el acopio, validación y análisis con herramientas de acceso gratuito	197
EJE 4: INFRAESTRUCTURA TECNOLÓGICA	221
La importancia de los Repositorios Institucionales para el seguimiento automático de los Objetivos de Desarrollo Sostenible	222
Reportes estadísticos para repositorios digitales a partir de múltiples fuentes basado en el <i>stack</i> ELK	234
Relevamiento de repositorios con perfiles de autor	248
PÓSTERS	267
Un sistema de índices para la relevancia de la producción grupal e individual	268
Evaluación de los repositorios institucionales del sector agrario y afines en el Perú: estudio preliminar	271
Representación de la colaboración de cuerpos académicos por medio de la teoría de grafos	274
Ampliação da ciência aberta na Universidade Federal de Minas Gerais: ações e perspectivas para ampla visibilidade dos resultados de pesquisa	277
Captação de recursos para projetos desenvolvidos pelo Núcleo de Sistema Integrado de Bibliotecas – NSIB do INCA: Lições aprendidas	292
Modelo de arquivamento de páginas web para Portais de Periódicos: uma iniciativa experimental do Portal de Periódicos da UNICAMP	295
Experiencia en la gestión de la productividad académica a través de colecciones dinámicas de LENS.ORG en la Universidad de Guadalajara, un estudio transversal	298
Repositorio Centroamericano de Patrimonio Cultural del IIRTE (UCR) como espacio para divulgar colecciones de obras utilizadas en investigación	302

OTRAS ACTIVIDADES _____	305
Consortio Iberoamericano para Educación en Ciencia y Tecnología (ISTEC). Nuevos proyectos _____	306
DiNaRe: El nuevo servicio de REMERI para la visibilidad del Acceso Abierto en México _____	308
AUTORES _____	311

Programa

BIREDIAL-ISTEC

Instituto Nacional de Astrofísica (INAOE), México

25 al 29 de octubre 2021

[Programa on line](#)

	Conferencia
	Ponencia
	Mesa
	Póster
	Otras actividades

Lunes 25 de octubre

8:00 **Proyección de video institucional del INAOE**

8:15 **Inauguración del evento**

EDMUNDO GUTIÉRREZ, Director General del INAOE

FRANCISCO RENERO, Director de Formación Académica, Comité Organizador Local

MARISA DE GIUSTI, Presidenta de ISTEC

MARLEM URIBE MARENCO, Comité Científico BIREDIAL-ISTEC

Enlace:



8:45 **Conferencia magistral**

Estado de situación y avances en ciencia abierta y acceso abierto en Argentina. Legislación, implementación, resultados, retos

A cargo de PAOLA AZRILEVICH, Coordinadora Biblioteca Electrónica de Ciencia y Tecnología, MinCyT (Argentina)

Modera: MARISA DE GIUSTI, Universidad Nacional de La Plata (Argentina)

Enlace:



9:45 **RECESO**

10:00 Enlace al bloque de ponencias:



Eje 1: Comunicación académica y científica en abierto

Modelo de gestión de ciencia abierta para la Universidad Autónoma de Zacatecas. Hacia la formación de investigadores en prácticas abiertas
A cargo de MONTSERRAT GARCÍA GUERRERO (México)

10:15 **Eje 1: Comunicación académica y científica en abierto**

Lineamientos institucionales sobre Ciencia Abierta: caso de la Universidad del Rosario en Colombia

A cargo de MARÍA LUCIA LIZARAZO RIVERO, MALGORZATA LISOWSKA NAVARRO y HUMBERTO BLANCO CASTILLO (Colombia)

10:30 **Sesión de preguntas**

Modera: MARLEM URIBE, Universidad del Norte (Colombia)

10:45 **Mesa de debate**

Políticas nacionales de Acceso Abierto en América Latina

Participantes:

VÍCTOR GÓMEZ RAZZA, Subdirector de Gestión de la Información y del Conocimiento, CONCYTEC (Perú)

PAOLA AZRILEVICH, Coordinadora Biblioteca Electrónica de Ciencia y Tecnología, MinCyT (Argentina)

BIANCA AMARO, Presidenta LA Referencia (Brasil)

Modera: MARISA DE GIUSTI, Universidad Nacional de La Plata (Argentina)

Enlace:



Martes 26 de octubre

8:00

Conferencia magistral

¿Cómo el Acceso Abierto diamante puede contribuir hacia una evaluación responsable e inclusiva de la ciencia?

A cargo de ARIANNA BECERRIL, Presidenta de AmeliCA (México)

Modera: MEILYN GARRO, Universidad de Costa Rica (Costa Rica)

Enlace:



9:00

Enlace al bloque de ponencias:



Eje 1: Comunicación académica y científica en abierto

Capacitação dos profissionais de informação visando a gestão no Arca. Repositório Institucional da Fundação Oswaldo Cruz

A cargo de CLAUDETE FERNANDES DE QUEIROZ (Brasil)

9:15

Eje 1: Comunicación académica y científica en abierto

Mapeamento dos projetos brasileiros em Ciência Cidadã e Biodiversidade

A cargo de LEILA MORÁS SILVA (Brasil)

9:30

Sesión de preguntas

Modera: CATERINA PAVÃO, Universidade Federal do Rio Grande do Sul (Brasil)

9:45

RECESO

10:00

Mesa de debate

Evaluación y métricas

Participantes:

CECILIA ROZENBLUM, Coordinadora General de Revistas Científicas, Universidad Nacional de La Plata (Argentina)

SÔNIA CAREGNATO, Profesora de la Facultad de Bibliotecología, Universidade Federal do Rio Grande do Sul (Brasil)

RICARDO VILLEGAS TOVAR, Coordinador de Producción Académica y Visibilidad Internacional Vicerrectoría de Investigación y Estudios de Posgrado, Benemérita Universidad Autónoma de Puebla (México)

ALEJANDRO URIBE TIRADO, Escuela Interamericana de Bibliotecología, Universidad de Antioquia (Colombia)

Modera: GONZALO VILLAREAL, Universidad Nacional de La Plata (Argentina)

Enlace:



11:00

Enlace al bloque de ponencias:



Eje 3: Evaluación y métricas alternativas

Datos bibliométricos para las Ciencias Sociales y las Humanidades

A cargo de SILVIA GUTIÉRREZ DE LA TORRE (México)

Modera: MEILYN GARRO, Universidad de Costa Rica (Costa Rica)

Miércoles 27 de octubre

8:15

Conferencia magistral

Datos Abiertos: Gestión de datos de investigación

A cargo de ENRIQUE SUCAR, INAOE (México)

Modera: Carlos Alberto Reyes García, INAOE (México)

Enlace:



9:15

Enlace al bloque de ponencias:



Eje 2: Datos abiertos

Colecciones públicas abiertas: propuesta de depósito digital de metadatos bibliográficos de las publicaciones oficiales de Bogotá

A cargo de WILSON ENRIQUE COLMENARES MORENO (Colombia)

9:30

Debate y preguntas

Modera: MALGORZATA LISOWSKA NAVARRO, Universidad del Rosario (Colombia)

9:45

RECESO

10:00

Mesa de debate

Datos de investigación en abierto

Participantes:

IANKO LÓPEZ ORTIZ DE ARTIÑANO, Director técnico, Consorcio Madroño (España)

HUMBERTO BLANCO, Universidad del Rosario (Colombia)

RAÚL SIFUENTES, Unidad de Automatización del Sistema de Bibliotecas, Pontificia Universidad Católica del Perú (Perú)

RAFAEL PORT DA ROCHA, Proyecto Red de Datos de Investigación Brasileña (RDP), Universidade Federal do Rio Grande do Sul (Brasil)

Moderadora: MALGORZATA LISOWSKA NAVARRO, Universidad del Rosario (Colombia)

Enlace:



11:00

Sesión de pósters

Captação de recursos para projetos desenvolvidos pelo Núcleo de Sistema Integrado de Bibliotecas – NSIB do INCA: Lições aprendidas.

ROBSON DIAS MARTINS, KÁTIA DE OLIVEIRA SIMÕES y CAMILA BELO FERREIRA

Repositorio Centroamericano de Patrimonio Cultural del IARTE (UCR) como espacio para divulgar colecciones de obras utilizadas en investigación. SOFÍA VINDAS SOLANO y MEILYN GARRO

Evaluación de los repositorios institucionales del sector agrario y afines en el Perú. SILVANA AQUINO REMIGIO

Representación de la colaboración de cuerpos académicos por medio de la teoría de grafos. JARED DAVID TADEO GUERRERO SOSA, MARÍA ENRIQUETA CASTELLANOS BOLAÑOS y VÍCTOR HUGO MENÉNDEZ DOMÍNGUEZ

Un sistema de índices para la relevancia de la producción grupal e individual. JARED DAVID TADEO GUERRERO SOSA, VÍCTOR HUGO MENÉNDEZ DOMÍNGUEZ y MARÍA ENRIQUETA CASTELLANOS BOLAÑOS

Ampliação da ciência aberta na Universidade Federal de Minas Gerais: ações e perspectivas para ampla visibilidade dos resultados de pesquisa. IZABEL ANTONINA DE ARAÚJO y CAMILA MARIANA APARECIDA DA SILVA

Modelo de arquivamento de páginas web para portais de periódicos. GILDENIR CAROLINO SANTOS y JULIANO CARDOZO

Experiencia en la gestión de la productividad académica a través de

colecciones dinámicas de LENS.ORG en la Universidad de Guadalajara,
un estudio transversal. NETZAHUALPILLI DELGADO FIGUEROA, GUADALUPE
JEANETTE GONZÁLEZ DÍAZ, ESMERALDA ALCARAZ SÁNCHEZ y MARTÍN
ADALBERTO TENA ESPINOZA DE LOS MONTEROS

Modera: LUIS ENRIQUE GARDUÑO, INAOE (México)

Enlace:



Jueves 28 de octubre

8:00

Conferencia magistral

Plataformas abiertas

A cargo de ELOY RODRIGUES, Confederation of Open Access Repositories
– COAR (Portugal)

Modera: FRANCISCO RENERO, INAOE (México)

Enlace:



9:00

RECESO

9:15

Enlace al bloque de ponencias:



Eje 4: Infraestructura tecnológica

La importancia de los Repositorios Institucionales para el seguimiento automático de los objetivos de desarrollo sostenible

A cargo de JOSÉ DANIEL TEXIER y JOHN ATKINSON (Argentina)

9:30

Eje 4: Infraestructura tecnológica

Reportes estadísticos para repositorios digitales a partir de múltiples fuentes basado en el stack ELK

A cargo de PABLO DE ALBUQUERQUE, MARISA DE GIUSTI y GONZALO LUJÁN VILLARREAL (Argentina)

9:45

Eje 4: Infraestructura tecnológica

Relevamiento de repositorios con perfiles de autor

A cargo de EZEQUIEL MANZUR, SANTIAGO TETTAMANTI, GONZALO LUJÁN

VILLARREAL, MARISA DE GIUSTI y ARIEL JORGE LIRA (Argentina)

10:00

Sesión de preguntas

Modera: ROSALINA VÁZQUEZ, Universidad Autónoma de San Luis Potosí (México)

10:30

RECESO

10:45

Mesa de debate

Infraestructura tecnológica: Charla técnica sobre infraestructuras tecnológicas para gestión de datos de investigación

Participantes:

LAUTARO MATAS, Secretario Ejecutivo y Técnico LA Referencia (Argentina)

LUIS ELIÉCER CADENAS, Director Ejecutivo de Red CLARA (Venezuela)

MANUELA KLANOVICZ FERREIRA, Analista de Sistemas del Centro de Procesamiento de Datos, Universidade Federal do Rio Grande do Sul (Brasil)

Modera: MARISA DE GIUSTI, Universidad Nacional de La Plata (Argentina)

Enlace:



Viernes 29 de octubre

8:00

Proyectos nuevos ISTEC

A cargo de MARISA DE GIUSTI, Presidenta de ISTEC (Argentina)

Enlace:



8:15

DiNaRe: El nuevo servicio de REMERI para la visibilidad del Acceso Abierto en México

A cargo de ROSALINA VÁZQUEZ y ANTONIO RAZO, Red Mexicana de Repositorios Institucionales – REMERI (México)

Enlace:



8:30

Mesa de debate

Experiencias de nodos nacionales de LA Referencia

Participantes:

WASHINGTON RIBEIRO, Instituto Brasileiro de Informação em Ciência e Tecnologia – IBICT (Brasil)

JUAN MALDINI, Gerente de Servicios Digitales en Agencia Nacional de Investigación e Innovación (ANII) (Uruguay)

PAOLA AZRILEVICH, Coordinadora Biblioteca Electrónica de Ciencia y Tecnología, MinCyT (Argentina)

ANDREA MORA, Sub Comisión de Conocimiento Abierto de CONARE (Costa Rica)

PAOLA CASTRO, Red Colombiana de Información Científica – REDCOL (Colombia)

FREDDY SUMBA, Especialista en desarrollo de servicios académicos de

CEDIA (Ecuador)

Modera: LAUTARO MATAS, Secretario Ejecutivo y Técnico de LA Referencia

Enlace:



9:45

RECESO

10:00

Premiación de pósters

Enlace:



10:15

Conclusiones del evento y Clausura

Enlace:



Presentación

Desde su fundación en 2011, el objetivo general de la conferencia internacional BIREDIAL-ISTEC ha sido promover la iniciativa de acceso abierto, la visibilidad institucional latinoamericana. y a partir de 2021, el evento adquirió un compromiso con la promoción de los demás componentes de la ciencia abierta.

La conferencia intenta reunir a las comunidades científicas, observatorios, bibliotecas, centros de datos abiertos, entre otros, y a los profesionales interesados en cuestiones relacionadas con la gestión, promoción y visibilidad del conocimiento así como a autoridades académicas, editores y directores de publicaciones científicas y alumnos de las distintas especialidades que tienen compromisos e intereses con la ciencia abierta.

La conferencia busca asimismo concitar la atención de las instituciones de enseñanza superior y centros de investigación, personal del gobierno y de la administración pública, interesados en compartir preocupaciones relacionadas con la transparencia de los datos públicos, para de este modo crear consenso sobre las mejores prácticas en las distintas áreas.

La edición 2021 se centró en cuatro ejes principales que sirvieron para impulsar los avances en ciencia abierta que se describen de manera abreviada aquí:

EJE 1: COMUNICACIÓN ACADÉMICA Y CIENTÍFICA EN ABIERTO

Gestión y modelos de sostenibilidad de repositorios

Estrategias para mejorar la calidad, visibilidad y posicionamiento de las revistas científicas, académicas o de divulgación

Modelos de gestión de Recursos Educativos Abiertos (REA)

Estrategias de visibilidad académica y científica en abierto

Estrategias de sensibilización, capacitación y formación de personal sobre AA y CA

EJE 2: DATOS ABIERTOS

Gestión de datos de investigación: planes, principios FAIR y CARE, derechos de autor

Gestión de repositorios de datos de investigación: curaduría de datos, normalización, datos enlazados, gestión de grandes volúmenes de datos, etc.

Minería de datos e inteligencia artificial: extracción, análisis e interpretación de datos, identificación de relaciones entre recursos, visualizaciones

EJE 3: EVALUACIÓN Y MÉTRICAS ALTERNATIVAS

Métricas alternativas para evaluar la actividad científica y su impacto. Evaluación de los distintos productos de la ciencia: artículos, revistas, datos, software, proyectos de investigación, presentaciones en congresos, etc. Actores de la Ciencia Abierta

Evaluación abierta por pares y metodologías abiertas

Herramientas, metodologías de extracción, y tratamiento de datos paraprocesos de evaluación (Google Scholar, Scopus, WoS, AmeliCA, Dimensions, Altmetric, DOAJ, etc.)

EJE 4: INFRAESTRUCTURA TECNOLÓGICA

Plataformas para la implementación de servicios de ciencia abierta: Repositorios de datos, de REA, Sistemas CRIS, libros electrónicos. Software libre o propietario

Interoperabilidad entre sistemas: CRIS, portales de revistas, repositorios de datos, repositorios de recursos educativos, ORCID, portales de visibilidad, etc.

Tecnología para la preservación digital de contenidos académicos y científicos

Tecnología para la extracción automática y análisis de datos

Nueva generación de plataformas abiertas

La décima edición de Biredial-ISTEC reunió distintas modalidades de presentación: pósteres, conferencias magistrales, mesas de debate y presentaciones de trabajos abarcando 68 autores de 13 países, que pueden apreciarse a continuación.

Conferencias magistrales (resúmenes)

Estado de situación y avances en ciencia abierta y acceso abierto en Argentina

Legislación, implementación, resultados, retos

PAOLA A. AZRILEVICH

Coordinadora Biblioteca Electrónica de Ciencia y Tecnología, MinCyT (Argentina)

Moderadora: **MARISA R. DE GIUSTI**

Universidad Nacional de La Plata (Argentina)

RESUMEN

En aras de garantizar que la información científica esté al alcance de todos, el Ministerio de Ciencia y Tecnología (MINCYT) de Argentina ha realizado y realiza diferentes acciones. En esta conferencia se dará un breve repaso sobre qué se ha hecho y qué falta hacer. Una de las acciones tomadas por el MINCYT es la creación de los sistemas nacionales: instrumentos que buscan alcanzar el uso eficiente de los equipamientos de gran porte y la mejor organización y acceso a los datos y publicaciones científicas de las instituciones nacionales; son redes de instituciones nacionales nucleadas a través del MINCYT, entre las que se cuentan: Microscopía (primer sistema creado en 2008), Rayos X, Resonancia Magnética, Espectrometría de Masas, Computación de Alto Desempeño, Láseres, Magnetometría y Citometría de Flujo. Otras redes incluyen Bioterios y Micro y Nanofabricación. En el mismo sentido, se realizan tareas de visibilidad y relevamiento, así como el fomento de repositorios digitales y archivos de documentación histórica, todo bajo el paraguas de la Ciencia Abierta (CA). Entre las temáticas todavía no abordadas se encuentran la investigación abierta y colaborativa, el software abierto, los Recursos Educativos Abiertos, las metodologías abiertas y la evaluación abierta (aunque existe ya una comisión y se está empezando a estudiar). En cuanto al Sistema

Nacional de Repositorios Digitales (SNRD), que tiene como propósito conformar una red interoperable de repositorios digitales en ciencia y tecnología, a partir del establecimiento de políticas, estándares y protocolos comunes, comenzó a funcionar de manera no oficial a partir del 2009 y fue oficialmente lanzado en 2011; en 2013 se lanzó el Portal SNRD (que ya tiene más de 300.000 objetos en acceso abierto) y se sancionó la Ley N° 26.899, reglamentada en 2016. En 2020 se lanzó DACyTAr, portal de datos abiertos, que permite acceder exclusivamente a conjuntos de datos primarios de investigación, que ya cuenta con 397 conjuntos de datos de 10 repositorios institucionales. Esto resalta la importancia de depositar los datos en los repositorios institucionales más allá de los repositorios temáticos y la importancia de la interoperabilidad (protocolo OAI-PMH) y de cumplir con las directrices de metadatos. En este sentido, próximamente habrá nuevas directrices, como OpenAIRE 4.0, que permitirán mayor riqueza en la descripción de los materiales y generar mejores indicadores. La ley 26.899 garantiza y exige que al menos las versiones de los autores de los trabajos estén disponibles en los repositorios institucionales, en tanto dichos trabajos han sido financiados por parte del Estado nacional, a través de salarios, incentivos, subsidios, y a través del acceso a bibliografía científica, infraestructura, insumos y otros. Se detallan en este sentido las obligaciones del ministerio, de las instituciones y de los autores. ¿Se cumple la ley? Comienza a cumplirse parcialmente: por ejemplo, de los 67 repositorios adheridos, sólo 49 tienen políticas de Acceso Abierto (AA) y de estos sólo 39 tienen depósito obligatorio y 23 exigen el depósito de datos, entre otros indicadores. Entre los problemas más frecuentes detectados en las políticas institucionales, se destaca que éstas son más bien reglamentos internos de los repositorios, a pesar de la existencia de una ley que las respalda. El MINCyT ofrece una herramienta de verificación para que las instituciones puedan comprobar si cumplen o no con la ley, así como financiamiento y capacitación en recursos humanos. Entre los asuntos pendientes sobre políticas institucionales, se destacan el desconocimiento de la ley, los reglamentos de becas que no incluyen cláusulas de depósito, bajo nivel de autoarchivo, «miedo» a abrir los datos, entre otros; en temas de

preservación, se observa ausencia de política de formatos y bajo nivel de adopción de formatos válidos, entre otros; en materia de interoperabilidad, siguen faltando identificadores persistentes para ítems y para personas, así como bases de datos de autores normalizados; en cuanto a infraestructura, se registran problemas de hardware, de back-up y de conectividad, desactualización de software; en cuanto a recursos humanos, debe insistirse en la multidisciplinariedad necesaria de los equipos para la administración de los repositorios. Cabe también destacar el hecho de que aún no hay una conciencia colectiva sobre las cuestiones de CA y AA, fundamentales para la soberanía informativa. En 2021, el MINCYT estableció el Comité Asesor en Ciencia Abierta y Ciudadana, con el objeto de brindar asesoramiento experto en el área y realizar un diagnóstico de la situación del país en cuanto a CA.

PALABRAS CLAVE

Acceso abierto; repositorios digitales; sistemas nacionales; ciencia abierta; políticas institucionales.

¿Cómo el Acceso Abierto diamante puede contribuir hacia una evaluación responsable e inclusiva de la ciencia?

ARIANNA BECERRIL GARCÍA

Presidenta de AmeliCA (México)

Moderadora: **MEILYN GARRO**

Universidad de Costa Rica (Costa Rica)

RESUMEN

Hay una necesidad ampliamente reconocida de transitar hacia modelos y estrategias de evaluación de la ciencia y la actividad investigadora donde se recurra a miradas más comprensivas de la comunicación y generación de conocimiento, que vayan más allá de métricas propietarias basadas en citación como el FI, en las cuales el Sur Global y el Acceso Abierto no comercial carecen de representación. Una filosofía concretada en la declaración de San Francisco DORA, apoyada por múltiples instituciones e iniciativas, entre las que se encuentran Redalyc, la cual desde el 2018 solicita a las revistas que indexa que manifiesten empatía por la causa de DORA. El Acceso Abierto no comercial o diamante es el que brinda el mayor potencial en la evolución de la evaluación de la ciencia, ya que es el único que no involucra restricciones ni exclusiones tácitas, tanto en el acceso y procesamiento de datos como en la participación transversal de la comunidad científica. Se presentan una serie de reflexiones y una propuesta metodológica que incluye el sentido de la noción de deconstrucción, como el acto de des-sedimentar, re-significar, re-politizar el «dato» (cargado de teoría), que a menudo ha servido para invisibilizar el trabajo de ciertas regiones, modelos de difusión del conocimiento o disciplinas como

las Ciencias Sociales y las Humanidades. La reflexión está encaminada a volcar y traducir dichos elementos en tecnología (software) y estructuras de contención de la información (bases de datos); así como a reorganizar lo semánticamente difuso (normalización), a unir la información básica (metadato en lenguaje máquina), bajo estructuras de interconexión (XML) y a potenciar la visualización (cartografías), para otorgarle un nuevo significado al trabajo académico en los sistemas de evaluación de la investigación.

PALABRAS CLAVE

Acceso abierto; acceso abierto diamante; Redalyc; América Latina; repositorios institucionales.

Datos Abiertos: Gestión de datos de investigación

ENRIQUE SUCAR SUCCAR

INAOE (México)

Modera: **CARLOS ALBERTO REYES GARCÍA**

INAOE (México)

RESUMEN

Las bases de datos (BD) son el «oro» de la investigación científica, por lo que es importante hacerlas accesibles a la comunidad; sin embargo, en algunos campos, como la medicina, es difícil encontrar bases de datos abiertas. No obstante, es posible «pre-procesar» estas BD para hacerlas accesibles y a su vez proteger los datos sensibles. Un buen ejemplo de esto es la Base de Datos Mexicana de COVID-19, resultado de la colaboración de la Secretaría de Salud y la Universidad Nacional Autónoma de México. Cuenta con 6,5 millones de individuos con 97 variables y se actualiza diariamente. Dicha BD está accesible para fines de investigación y presenta información de carácter público recopilada a través de la Dirección General de Epidemiología de la Secretaría de Salud, que es curada y divulgada por la UNAM. LA BD incluye información de pruebas, hospitalización y decesos de 5.186 unidades públicas y privadas distribuidas en los tres niveles del sistema de salud mexicano. El proceso de curación de esta BD consistió, en primer lugar, en la eliminación de toda la información sensible (especialmente datos de identidad de los individuos), por parte de un equipo de expertos de la UNAM. Luego, se realizaron distintas tareas de estandarización de los datos, que facilitaron los procesos de minería de datos y otros que la vuelven accesible con fines de investigación. La BD

accesible contiene 97 campos que pueden agruparse en Información general del paciente, Síntomas, Comorbilidades, Diagnóstico y tratamiento y Resultados. Para el análisis se seleccionaron 47 variables, a partir de las cuales se pueden obtener estadísticas (se muestran ejemplos con distintos gráficos): evolución de la pandemia; número de casos por ocupación; número de casos por síntomas; recuperados y defunciones por comorbilidades; riesgo de fallecimiento por edad y comorbilidades. Modelos causales: en las técnicas tradicionales de minería de datos se obtienen asociaciones o correlaciones entre las variables que pueden ser engañosas; para evitar esto e ir más allá, es necesario aprender sobre relaciones causales, lo que se conoce como «descubrimiento causal». Se trata del co-factor (o causa común) que puede haber entre dos variables y que explicaría la correlación encontrada. En general, los humanos pensamos en términos causales (tal cosa causa tal otra), por lo que la ciencia de la causalidad procura entender este mecanismo de razonamiento causal formalmente, para emularlo luego en las computadoras. En los últimos años ha habido importantes avances en este sentido, especialmente en los modelos gráficos causales: representación del conocimiento causal e inferencia causal (poder predecir el efecto de determinadas intervenciones o imaginar escenarios alternativos). Hay tres niveles de causalidad: asociación (regularidad); intervención (predicción de efectos) y contrafactuales (imaginación, introspección). Los modelos estadísticos tradicionales sólo permiten hacer asociaciones, mientras que los modelos causales permiten la intervención y los escenarios contrafactuales. En este sentido, se destacan las redes bayesianas causales, que son modelos gráficos que representan variables causales (arcos) entre variables (nodos) e implican suposiciones más fuertes que las redes bayesianas y permiten hacer razonamientos causales. Para aprender entonces relaciones causales es preciso hacer experimentos; por ejemplo, lo que se ha hecho con las vacunas contra el COVID; pero no siempre es posible hacer estos experimentos. En cambio, es posible obtener modelos causales a partir de datos observacionales, que abundan en muchos dominios. Para ello, se debe incluir una serie de suposiciones (suficiencia causal, por ejemplo), asumir cierto tipo

de distribuciones de los datos (modelos lineales gaussianos), tener un gran conocimiento previo de los datos y realizar algunas intervenciones para distinguir entre modelos equivalentes, estadísticamente hablando. Para esto último, se utiliza el cálculo de efectos causales, en el que se interviene una variable y se estima el efecto en otras. Actualmente, se está tratando de encontrar información causal en la Base de Datos Mexicana de COVID-19, de modo que pueda servir para ayudar a los investigadores. Para ello, se aplicaron diversos algoritmos de aprendizaje causal en la BD, se enfocó en la relación de ciertas variables con la mortalidad y se analizaron los datos de las tres principales olas de la pandemia en México para tratar en entender mejor el fenómeno y las diferencias entre las etapas. Se muestran algunos resultados para las ciudades de Yucatán y CDMX. Esta investigación, todavía en curso, demuestra que la BD es un buen ejemplo de lo que puede obtenerse con estas técnicas, que permitirán desarrollar sistemas inteligentes más robustos y explicables, así como generar conocimiento útil para la toma de decisiones. Para el trabajo a futuro, se continuará analizando la BD, se incluirá información sobre la genética del virus y se realizarán nuevos modelos de sujetos y de grupos específicos para poder hacer una mejor transferencia de conocimiento.

PALABRAS CLAVE

Bases de datos; datos abiertos; COVID-19; modelos causales; descubrimiento causal.

Repositorios, plataformas abiertas y comunitarias: para la sostenibilidad, innovación y equidad en la comunicación científica y académica

ELOY RODRIGUES

Confederation of Open Access Repositories – COAR (Portugal)

Moderador: **FRANCISCO RENERO**

INAOE (México)

RESUMEN

El sistema de comunicación científica actual no es abierto, sostenible, inclusivo o innovador. Por ejemplo, en cuanto a los costos, según datos de STM Global Brief 2021, el mercado editorial global mueve unos 27 billones de dólares por año (10,8 de revistas, 3,2 de libros y el resto bases de datos, plataformas y otros productos). Se estima un ingreso de 4.000 a 5.000 euros por artículo publicado en revistas de suscripción paga y el valor medio de los APC es de 2.000 euros (aunque ha venido aumentando en los últimos 5 años). ¿Cuáles son las funciones de las revistas científicas? Registro (reconocimiento de la primacía intelectual); certificación (calidad/validez de una investigación); divulgación (accesibilidad y difusión) y archivo (preservación de la investigación para uso futuro). ¿Qué es necesario para cambiar? Nuevos modelos de publicación y comunicación y nuevas prácticas de evaluación de la investigación. Los nuevos modelos se han visto golpeados por la disfuncionalidad del sistema de comunicación académica, la inflación de los precios y el aumento del control por parte de los proveedores, el aumento de los preprints, los estímulos de la política a la investigación abierta y, desde

luego, la pandemia por COVID-19. Esto último, sin embargo, ha significado un enorme crecimiento para las prácticas de ciencia abierta (datos abiertos, acceso abierto a las publicaciones, publicación de preprints, revisión rápida y revisión abierta, publicación «overlay»). Durante la pandemia, por ejemplo, el enfoque se puso en la colaboración y la compartición para el impacto social y hubo un uso de canales y modelos innovadores de publicación y difusión. Muchos expertos han destacado que este puede ser un punto de inflexión y se puede propiciar un cambio de paradigma en la comunicación científica, aprovechando esta circunstancia fortuita. Es posible pensar en una «disociación» de las funciones de la publicación científica, en la que la certificación y divulgación vayan por un lado y el registro y el archivo por otro, específicamente en estos últimos dos aspectos entran a jugar fuertemente los repositorios (Pubfair, Épisciences). Lo que se propicia es cambiar el paradigma de «revisar y después publicar» a «hacer público y revisar». Desde luego, esto necesita la aceptación y adopción por parte de la comunidad y una infraestructura acorde, abierta, sostenible, actualizada y gobernada por la misma comunidad. En este sentido, la visión y el trabajo de COAR consisten en reforzar y expandir el papel de los repositorios y sus comunidades, en la gestión de todos los resultados de investigación. *Next Generation Repositories*: necesitamos modernizar y mejorar los repositorios existentes. COAR Notify Project: este proyecto procura experimentar con servicios de revisión por pares «overlay». La idea es aprovechar, mejorar y añadir valor a la actual inversión institucional y de los financiadores en plataformas abiertas para transformar así el sistema de comunicación científica.

PALABRAS CLAVE

Acceso abierto; ciencia abierta; plataformas; repositorios institucionales; publicación «overlay».

Mesas de debate
(transcripciones)

Mesa 1: Políticas nacionales de Acceso Abierto en América Latina

BIANCA AMARO

Presidenta de LA Referencia (Brasil)

PAOLA A. AZRILEVICH

Coordinadora de la Biblioteca Electrónica de Ciencia y Tecnología, MinCyT (Argentina)

VÍCTOR H. GÓMEZ RAZZA

Subdirector de Gestión de la Información y del Conocimiento, CONCYTEC (Perú)

Modera: **MARISA R. DE GIUSTI**

Universidad Nacional de La Plata (Argentina)

RESUMEN EXTENDIDO

¿Cuáles cree que son las fortalezas que se destacan de la política de acceso abierto de su país y cuáles son las debilidades? Para el caso específico de Brasil, ¿cuánto suple, en este sentido, el mandato de FAPESP, y qué destaca de ese mandato que podría incluirse en una ley nacional?

BIANCA AMARO (Brasil): El mandato de FAPESP puede ayudar y ser una base para todas las agencias de financiación; intenta arreglar de algún modo cómo deben comportarse los proyectos financiados por FAPESP. En mi experiencia, y mirando todas las políticas de acceso abierto en la región, es envidiable la situación de Argentina, que tiene su ley sancionada. Debemos ser más rigurosos incluso y nuestras leyes deben ser mandatorias. Las leyes que traen recomendaciones están muy bien y sobre todo en aquel momento, en el que la ciencia abierta no estaba tan sólida y reconocida como ahora. Por eso, en una ley brasileña, buscaría primero cuestiones mandatorias, ya que como

funcionarios públicos toda nuestra actividad está pagada con fondos públicos; en segundo lugar, que haya un plan de gestión de datos con formatos internacionales, adecuado a cada disciplina, desde luego. Rogaría además que se sigan las actividades de la Research Data Alliance (RDA), pues allí se discuten las normativas internacionales relacionadas con los datos de investigación.

PAOLA AZRILEVICH (Argentina): Además de que me llena de orgullo que Argentina tenga su ley de acceso abierto, es fundamental que sea obligatoria, pues de lo contrario no sería una ley; otro de los aspectos fundamentales es que se adopte la vía verde como política de Estado para el acceso abierto; asimismo, no obliga a nadie a publicar únicamente en revistas de AA y pagar costos de APC, sino que permite que cada uno publique y difunda su producción científica por las vías que quiera en una primera instancia, obligando, en una segunda instancia, con plazos claros, cortos y establecidos por la ley, lo que también es una fortaleza, a difundir en los repositorios de sus instituciones. La vía verde como política de Estado obliga a las instituciones a que gestionen su producción científica y el modo de difundirla, que era algo de lo que las instituciones, en general, no se hacían muy responsables antes: los investigadores por un lado, la institución por otro, y cada tanto algún informe. La interoperabilidad es otra cuestión central en nuestra ley y la posibilidad de integrar sistemas nacionales, regionales e internacionales no es, sin dudas, algo menor. Las instituciones tienen obligación, por ser parte del Estado y estar financiadas por el Estado, de tener todo lo que producen accesible para la comunidad en su totalidad. Otra fortaleza de la ley es que tiene excepciones claras y que además excluye las cesiones de derechos a las editoriales y ese es un derecho importantísimo que da esta ley, así como una garantía. La soberanía informativa es otro de sus pilares, pues tenemos derecho como país a ser reconocidos por todo lo que se produce y almacena en nuestros repositorios. Como debilidad, se puede mencionar las sanciones, pues si bien por un lado puede tomarse como una fortaleza el hecho de que haya sanciones si no se cumple la ley, por otro lado se da el hecho de que muchos creen que no se aplicarán estas sanciones. Faltan, en ese sentido, algunas instancias

intermedias. Otra debilidad que se puede mencionar, no relacionada con la ley, sino como organismo de control, es la capacidad de monitoreo. Necesitamos implementar y fortalecer esta capacidad de monitoreo para lograr que se cumpla la ley.

VÍCTOR GOMEZ RAZZA (Perú): Uno de nuestros orgullos es que hemos sido el primer país de América Latina en aprobar una legislación nacional de acceso abierto, con la creación de nuestro repositorio nacional en 2013. Según esta ley, todas las entidades y personas que componen el sistema nacional de CTI pueden registrar su producción en sus propios repositorios y CONCYTEC puede recolectar esta información para el repositorio nacional. Este ente recolector, ALICIA (Acceso Libre a Información Científica para Innovación), es el repositorio nacional y establece la obligatoriedad de publicar todos los resultados de investigaciones financiadas con fondos públicos. La ley define además al repositorio nacional como el ente centralizado digital, por lo que apoya normativamente el trabajo en acceso abierto en el país. Otra de las orientaciones que tiene la política de acceso abierto en Perú, es la adopción de estándares internacionales. Se está trabajando en ese sentido y se tienen ya normativas y directrices para que todas las entidades «hablen el mismo idioma» y se pueda interoperar. Tenemos también una importante presencia a nivel regional; tenemos 167 repositorios institucionales peruanos registrados en OpenDOAR, y tenemos 9 mandatos y políticas de acceso abierto registrados en ROARMap. Nuestra comunidad de acceso abierto se puede considerar consolidada, en el sentido de que tenemos una red, también en el marco de la ley, la Red Nacional de Repositorios Digitales de Ciencia, Tecnología e Innovación de Acceso Abierto, conformada por las instituciones que se han adherido a ALICIA. A la fecha ya hay 226 repositorios integrados. En cuanto a las debilidades, falta todavía fortalecer las capacidades funcionales y técnicas con respecto a la gestión de la información en acceso abierto y, tal como mencionaba Paola, un mayor seguimiento y monitoreo para velar porque se cumplan los objetivos buscados. Actualmente no estamos tampoco registrando toda la información referida a los proyectos por lo que no se pueden elaborar estadísticas al respecto.

¿Cuánto cree que se ha podido cumplir o avanzar en la política de acceso abierto y cuáles son los aspectos incumplidos u obturados y las dificultades y por qué?

B. AMARO: El hecho de no tener una política de AA en Brasil nos complica la acción, pero de todos modos hemos creado la Red Brasileña de Repositorios Digitales, formada a su vez por redes regionales, y así podemos tener una acción conjunta en estos términos. Mientras no tengamos una ley, vamos «inventado» estrategias para funcionar sin ella. Hemos estado así diseminando la idea del AA y ahora estamos empezando con la Ciencia Abierta. Hoy día tenemos también el Portal Brasileiro de Publicações e Dados Científicos em Acesso Aberto (Oasisbr), que tiene más de 2.700.000 documentos, lo que es un número enorme, pero no es tan grande como debería ser, dado el tamaño de nuestro país. Pero algo que resulta esencial y que debemos hacerlo ya, pero que es muy complejo y eso impide que la cosa avance, es la cuestión de las evaluaciones, es decir, cómo evaluamos a nuestros científicos. Tenemos que cambiar nuestros modelos de evaluación. En este sentido, el IBICT ha creado, dentro de la Alianza para el Gobierno Abierto, un compromiso para cambiar los mecanismos de evaluación; por el momento cuenta con varios actores y al menos hemos empezado a discutirlos, porque es preciso presentar un modelo, un modelo que se base en la CA. La recomendación de Unesco será muy importante en este sentido, para poder actuar y cambiar las cosas. Pero mientras los investigadores no vean que esto será bueno para ellos y que tendrán ventajas con este nuevo modelo, no lograremos muchos cambios.

P. AZRILEVICH: Siguiendo con la línea de Bianca, en Argentina, si bien con ley y todo, los modelos tradicionales de evaluación no son un motivo válido para el no cumplimiento de la ley; claramente los cambios, hacia modelos de evaluación abiertos, también van a movilizar mucho más rápido cualquier cosa que hagamos por vías legales. Sería importante, no sólo para Brasil sino para Argentina y para toda la región y el mundo en general, empezar a mirar otros indicadores a la hora de evaluar. En Argentina tenemos 139 instituciones alcanzadas por la ley de acceso abierto, de las cuales sólo 55 forman parte del Sistema Nacional de Repositorios Digitales (SNRD); es decir que todavía hay muchas instituciones que ni siquiera se han adherido con un proyecto de

repositorio o un borrador de política institucional. Tenemos 67 repositorios adheridos, esto es porque algunas instituciones tienen más de un repositorio, de los cuales están siendo cosechados 44, es decir, que cumplen con la normativa de interoperabilidad, y sólo 10 repositorios que ya tienen conjuntos de datos, pero lo cierto es que por lo menos los 44 cosechados deberían tener ya algún conjunto de datos, porque la ley, vigente desde el 2013, dice que hay un plazo de cinco años desde que los datos se generan para que estén disponibles en AA en los RI. Y por el momento tenemos sólo un plan de gestión de datos implementado por una institución, lo cual es realmente bajo porque cada institución debería tenerlo, o al menos un modelo genérico de PGD que les permita empezar a trabajar en ese sentido. Es preciso que empiecen a trabajar pensando que los datos deben ser gestionados de algún modo y si esto no se plantea de algún modo tampoco va a suceder. Incluso es difícil para sus usuarios, pues deben acceder a conjuntos de datos sin ningún contexto, cuando el PGD les proporcionaría esa información. La ley también dice que las políticas de AA deben hacerse en el marco de la ley y a nivel institucional; no deben ser el «reglamento del repositorio», ni un «mandato voluntario» u «optativo»; tiene que ser algo obligatorio, establecido por la institución con sus reglas de juego y con el piso que nosotros ponemos como ley. Lo que esperamos es que las instituciones crezcan por sobre ese piso, pues no les ponemos un «techo». Tenemos 49 políticas registradas, de las cuales no todas están aprobadas, muchas están en instancia de borrador todavía, y, por ejemplo, sólo 23 exigen el depósito de datos cuando es algo obligatorio por la ley. Y sólo 11 cumplen con el plazo de los seis meses cuando la ley es bien clara al respecto y especifica que tras ese plazo la publicación tiene que estar disponible en AA. Y otra cuestión que me parece aún más sensible y grave, es que muchas instituciones todavía son «permeables» a las decisiones de las editoriales, aun teniendo una ley nacional que las respalda. Creo que en este sentido nos falta un trabajo de marketing bien entendido, bien hecho y muy agresivo hacia las instituciones para que tomen conciencia en torno a esto, para después poder exigir a la comunidad científica. Si las instituciones no asumen su rol y no comprenden la importancia de esta situación, es muy difícil

pedirle a la comunidad científica que pueda hacer su parte (no tienen las políticas institucionales que les dan el marco local, no tienen la infraestructura necesaria, no tienen asesoramiento) y así quedan a la deriva, a la voluntad de los derechos que cedieron, etc. e incluso muchas veces ignorando que tienen un repositorio en su propia institución. Vale decir que la falencia principal que estamos teniendo tiene que ver con la falta de una buena estrategia comunicacional.

V. GÓMEZ RAZZA: En Perú también se repiten algunas de las cuestiones que mencionaron tanto Bianca como Paola. Por ejemplo, una de las prioridades de gestión actualmente es mejorar y fortalecer las políticas de AA. A pesar de que tenemos casi el 90 % de instituciones adheridas a ALICIA, tenemos una falencia en la adopción de handles y DOI, y ahí tenemos un trabajo permanente de monitoreo y soporte, porque muchas veces es por falta de capacidades técnicas o alguna otra falencia que existe en la institución, dado que cada una tiene una realidad distinta, así que hay que estar acompañando y muchas veces en nuestros eventos tomamos estas distintas realidades y las combinamos para ver qué tienen en común y cómo podemos ayudarlos. Estamos también priorizando el registro de repositorios en OpenDOAR, que cumplan con los requisitos de calidad requeridos (también estamos ayudando a quienes todavía no los cumplen), y estamos apoyando fuertemente el depósito de post-prints y versiones finales en repositorios institucionales en cumplimiento de la ley. Siempre estamos en contacto permanente además con LA Referencia, que son muy amables en brindarnos su apoyo, y hemos estado participando en distintos eventos relacionados con ellos y estamos viendo la posibilidad de promover la implementación de un Directorio Latinoamericano de Derechos de Explotación y permisos para el autoarchivo, algo similar a Dulcinea o Sherpa-Romeo, que es algo que está faltando actualmente. Otro punto importante, en coincidencia con Paola y Bianca, es mejorar los criterios de evaluación de la calidad e implementación de repositorios en relación a las políticas de AA. En este tema tan cambiante es preciso hacer un seguimiento e ir afinándolos poco a poco. Otro punto importante es que una correcta implementación de repositorios de

datos de investigación debe seguir una mayor cobertura de las publicaciones y proyectos asociados.

Aprovechando que en esta mesa hay participantes de tres países diferentes (Argentina, Brasil y Perú), ¿consideran posible conciliar un avance coordinado en América Latina, especialmente a la luz de los esfuerzos necesarios para alcanzar las premisas de la ciencia abierta? ¿Cuáles son los aspectos que consideran débiles en las políticas de Latinoamérica?

B. AMARO: Por supuesto que creo que es posible; LA Referencia está para esto y nuestra tarea dentro de la red es apoyar las buenas prácticas, saber quién está desarrollando qué, y pensar en conjunto. Y también tener asiento y voz en las principales iniciativas internacionales, para que podamos decirles cómo es que las cosas funcionan por aquí. Porque los países del Norte creen que son los dueños del mundo y de hecho tenemos que acabar con esta visión de que tenemos que someternos a todo lo que viene del Norte. Esto para mí es clave. Incluso tuve una discusión dentro de este grupo en el que estamos discutiendo la evaluación, que era que tenemos que «esperar» que los del Norte... y yo «no, no tenemos que esperar, es nuestra oportunidad de que enseñemos al mundo nuestra manera de actuar». Porque somos fuertes, tenemos la cuestión del AA diseminada por toda la región y por lo tanto tenemos capacidades para esto. Mi idea siempre ha sido que juntos somos más fuertes, entonces tenemos que discutir todas estas cuestiones, qué funciona y qué no en esta política o en aquella, cómo podríamos perfeccionarla todos en conjunto. Por ejemplo, a mí me interesa conocer cómo funcionan las políticas de Argentina y de Perú, porque cuando logre una política para Brasil tendré en cuenta los problemas y qué habría que ajustar o no. Esta es la única ventaja de no tener todavía una política. Así que yo creo que trabajar juntos, discutir juntos es lo que debe hacerse, porque en este sentido, en la CA, somos todos iguales, no hay competencia. Tenemos que trabajar en la divulgación, pero fundamentalmente en la «recompensa» y la recompensa es, en mi opinión, una mejor evaluación de los investigadores. Tenemos que trabajar y pensar

todos juntos y aprovechar las mejores ideas de los unos y de los otros. O sea, LA Referencia es la respuesta.

P. AZRILEVICH: Me parece que esa es la apuesta que tenemos que hacer, que LA Referencia es el lugar que hemos sabido construir y sostener, lo que no es menor. Es un organismo resultado del esfuerzo de los Estados, de una decisión de gobierno de cada uno de nuestros países, no es una red conformada por voluntades de personas o grupos que se juntan porque tienen ganas de hacer algo, sino que además hay un compromiso de los países. Me parece que esto es vital y que ese bloque regional que conformamos con LA Referencia y que nos permite posicionarnos ante el resto del mundo con voz y voto en muchos lugares, es lo que nos permite pensar que vamos a poder cambiar algunas cosas. Así que, sí, LA Referencia es la respuesta, sin lugar a dudas.

V. GÓMEZ RAZZA: De acuerdo con Bianca y Paola, como miembros activos de LA Referencia también considero que juntos somos más y vamos a poder hacer un mejor trabajo, sobre todo con respecto a las actividades que realizamos aquí en Perú: son distintas realidades que juntamos, a nivel local, y lo mismo se repite en LA Referencia, a nivel de países, por lo que cada experiencia, por más que falte mucho trabajo por hacer, igual sirve. Es además importante el hecho de que LA Referencia sea la respuesta, como estamos diciendo, y que además tengamos esta entidad que nos agrupa y nos permite alinear objetivos y alinear también esfuerzos. Hace un momento, Bianca mencionaba la recomendación de la Unesco sobre CA y me parece muy importante, porque fortalece la búsqueda de un consenso mundial y si bien aparentemente estamos un poquito relegados, igual es importante que haya un consenso y transformar desde las agencias de gobierno los modelos de evaluación de los investigadores. Ese es un punto clave, que ya lo hemos mencionado. Básicamente, considero que la clave está en mantener a LA Referencia como la respuesta y apoyarla en los esfuerzos que viene realizando.

Mesa 2: Evaluación y métricas

CECILIA ROZENBLUM

Coordinadora General de Revistas Científicas, Universidad Nacional de La Plata (Argentina)

SÔNIA E. CAREGNATO

Facultad de Bibliotecología, Universidade Federal do Rio Grande do Sul (Brasil)

RICARDO VILLEGAS TOVAR

Coordinador de Producción Académica y Visibilidad Internacional Vicerrectoría de Investigación y Estudios de Posgrado, Benemérita Universidad Autónoma de Puebla (México)

ALEJANDRO URIBE TIRADO

Escuela Interamericana de Bibliotecología, Universidad de Antioquia (Colombia)

Modera: **GONZALO L. VILLARREAL**

Universidad Nacional de La Plata (Argentina)

RESUMEN EXTENDIDO

¿Cuáles son los nuevos indicadores, entre todos los indicadores que están apareciendo, que consideran más relevantes?

SÔNIA CAREGNATO (Brasil): En primer lugar, quiero dejar claro que, para mí, el principal indicador para evaluar trabajos científicos es el número de citas recibidas por el trabajo. No estoy hablando del factor de impacto, ni otros indicadores de impacto como Cite Score, basado en Scopus, sino que me refiero al número de veces que el trabajo aparece en la lista de referencias de otro trabajo científico. En esos casos, las citas muestran un uso o un efecto de una determinada investigación en la literatura científica del área. Así, las citas son reflejo de las referencias hechas por otro científico, de modo que trabajos anteriores son todavía influyentes e importantes. ¿Qué se mide entonces

cuando se mide el número de citas de un trabajo? El uso de esos trabajos por parte de la comunidad científica y, por tanto, aquello que nosotros llamamos impacto científico o impacto de citas de los trabajos. ¿Cuál sería la novedad de este indicador? No hay ninguna novedad, ya que puede ser elaborado a partir de distintas bases de datos. La novedad serían las fuentes de información o las herramientas que permiten calcular este indicador de citación. Sólo a modo de recordatorio, las tres herramientas más utilizadas son Web of Science (WoS), Scopus y Google Scholar (GS), más recientemente. Las críticas a los indicadores más tradicionales se dirigen principalmente al hecho de que son derivados de estas herramientas. Por un lado, porque tanto WoS como Scopus son muy selectivas en la indexación de las revistas, haciendo que buena parte de la literatura científica del mundo sea ignorada por ellas y, por otro lado, se critica también la falta de calidad de los metadatos indexados por Google Scholar y asimismo la facilidad de manipular los datos en GS. Estas críticas pueden ser válidas para pensar nuevas herramientas de validación. Como nuevos indicadores, podemos mencionar aquellos que buscan evaluar el impacto social de las investigaciones. Las citas miden el impacto científico, académico, y ahora estamos buscando cada vez más métricas para evaluar el impacto social de las investigaciones. Las altmetrics prometen esa evaluación y tal vez también se pueda medir el impacto en los medios sociales. Hay varios de estos nuevos indicadores que han surgido y uno bien interesante me parece que es el número de lecturas en Mendeley, que es un indicador calculado sobre el número de usuarios que guardaron un archivo específico en su biblioteca de esa red social. Ese indicador puede servir para evaluar la actividad científica consumida por otros investigadores y hasta por estudiantes, por lo que considero que es un indicador muy interesante.

CECILIA ROZENBLUM (Argentina): Elegir dos o tres indicadores me parece que es realmente un desafío. Creo que los indicadores de uso, de descarga y de cómo se maneja el contenido de las revistas en la web es fundamental. Necesitamos indicadores que nos digan qué está pasando una vez que publicamos un artículo, es decir, qué pasa después de que lo publicamos, pensando además en la inmediatez de la publicación. Es decir, inmediatamente

se comparte en las redes sociales, se descarga, se incluye en Mendeley u otros sitios similares, por lo que esa inmediatez, que no da la citación, me parece muy importante a la hora de pensar indicadores de evaluación. Porque ese es el primer impacto, la primera visibilidad que tiene el artículo o el contenido que está en línea. En este sentido, también me parece que ahí se evalúa la calidad editorial de las revistas científicas o el esfuerzo editorial de qué es lo que hace el equipo editorial para que el material esté más visible rápidamente. O sea, cómo se hace para que lo coseche un buscador o cómo se hace para que eso rápidamente corra por las redes sociales o si la revista tiene un *community manager* o algo por el estilo para que eso se difunda, etc. Ahí se evalúa no solamente el contenido y el artículo sino el esfuerzo de la revista por hacer no solamente que tenga un proceso editorial óptimo, sino que una vez que se publica eso sea difundido de la mejor forma posible, por la mayor cantidad de canales posible. Porque, claro, podemos asegurar un contenido perfecto pero que puede quedar «estancado» en el OJS de la revista si no se lo difunde y eso no sirve para nada. Los indicadores de uso, entonces, unidos a la analítica web de la revista son excelentes indicadores, sobre todo del uso más inmediato. La citación, la colaboración son otro tipo de indicadores, más de mediano plazo, complementarios de estos indicadores de este «primer impacto» de los artículos, sobre todo en estos tiempos donde hay tanto contenido en línea. Un detalle más es el de los formatos de lectura: ese otro esfuerzo editorial es muy interesante para pensar en las nuevas estadísticas que están saliendo de cómo los lectores están usando los contenidos desde distintos dispositivos.

ALEJANDRO URIBE TIRADO (Colombia): Lo que necesitamos es una mirada integral, holística, volver a las preguntas originales: ¿para qué investigamos? ¿para qué publicamos? ¿a qué público debe llegar la investigación? Creo que eso es lo que se ha olvidado, porque los sistemas de métricas de las universidades de nuestros países nos están pidiendo que miremos solamente un público o un ángulo. Yo creo que los indicadores son como tratar de tomar una foto panorámica, pensando que lo que se publica e investiga es para que llegue realmente a alguien. Entonces, estoy de acuerdo con Sônia: no es que la citación no sirva sino que sólo nos da un ángulo de la foto y para un

determinado sector del público. El problema de la citación ha sido que unas empresas se han apropiado de ese indicador: es decir, sirve la citación pero si está en WoS o está en Scopus y ellas son sólo unas fuentes determinadas en las que gran parte de la producción científica del mundo no está, una gran parte de citas no se recogen y por eso hay un movimiento del «open citation». Es decir, ¿qué nos dan las citas? Se publica un trabajo, hay un público especializado que se llama «investigadores», al que ese trabajo le sirve para generar nuevo conocimiento. ¿Eso sigue siendo válido? Sí. En ese sentido, a mí me gusta más Google Scholar porque da una visión más amplia de la citación, aunque tenga problemas de normalización, pero cuando se encuentra un trabajo en GS y dice «diez citas» o x cantidad de citas y se las revisa, uno puede ver que, por ejemplo, cuatro citas son de artículos de otros colegas (eso es impacto científico y está muy bien); dos de revistas que estaban en WoS o en Scopus y dos de revistas que no están allí, lo que indica otros públicos. Otras cuatro citas pueden de tesis, dos de pregrado y dos de posgrado, lo que está diciendo que ese artículo sirve para que los estudiantes se formen o para que un profesor pueda enseñar un contenido, con lo que está impactando en otros públicos, más desde la docencia que desde la investigación, y de pronto, otras dos citas, una es de una referencia de una patente (y entonces el artículo sirvió para que una organización hiciera algún tipo de innovación empresarial o tecnológica) y la otra proviene de una política pública. Eso lo da solamente GS en el mundo, porque recoge toda la citación posible; WoS y Scopus sólo están diciendo dos citas de esas diez porque son las que estaban en sus bases. Entonces, la citación no está mal, sino que el problema son las fuentes que utilizamos y que no se tiene en cuenta que un artículo científico no sólo cita otros artículos científicos sino que puede citar libros, capítulos de libros, tesis, patentes, políticas públicas, proyectos, etc. La citación sirve porque dice «a alguien le sirve este trabajo» para referenciarse y tomar algún elemento, sea teórico o práctico. Por eso debemos ampliar las fuentes y ver este otro tipo de citaciones, que no las estamos contando y ver qué nos dicen en cuanto al impacto científico y social. Creo que por eso las altmetrics empezaron a ser llamativas, pues miden muchas cosas. Parte de eso es visibilidad, no es

citación; y dependiendo de la herramienta de altmetrics se pueden mirar patentes, otras pueden mirar si los trabajos están citados en el currículo o en los planes de estudio, lo que es muy importante. Hay trabajos que se hacen para que sirvan para la docencia: ¿eso está mal? No. Cada trabajo tiene su alcance y su público. El éxito de las altmetrics es que nos están regalando una aproximación a la visibilidad o al impacto en otras fuentes, más allá de la citación tradicional. Asimismo creo que es importante no olvidar que no todas las cosas están registradas. Hay cosas que pueden estar sirviendo pero que no están registradas y ahí llega lo cualitativo. Creo que eso se ha perdido y eso se logra capturar en eventos, en proyectos, porque de pronto no salen en WoS y entonces no los pueden citar o ni siquiera Google Scholar los logra encontrar. Sin embargo, por ejemplo una maestra puede encontrar un trabajo de educación en acceso abierto (por eso hay que defender el acceso abierto también) y con eso sus alumnos leen mejor o la maestra hizo alguna innovación educativa. Esa maestra nunca va a publicar lo que hizo, sencillamente lo está aplicando en su práctica pero cuando uno capta esa información que está tácita o que está en otros contextos, eso está diciendo que ese trabajo sirvió. Así que habría que tratar de mirar cómo capturamos ese tipo de información «perdida», esa «citación perdida». Así, tendríamos tres categorías de citación: la tradicional, a la que hay que ampliarle las fuentes, las altmetrics, que tienen mucho potencial aunque miden cosas distintas, y luego la evaluación, que existe en muchos campos pero que todavía no lo medimos y el reto sería tratar de medirlo. Hay un gran campo por explorar. Hay que salir de la tradición del factor de impacto y del índice h y enseñarle a nuestros profesores y especialmente a los directivos y a los organismos de ciencia que existen otros indicadores, más allá de los que conocemos por los monopolios.

RICARDO VILLEGAS TOVAR (México): Siguiendo lo que mencionaba Alejandro hace un momento, quisiera comentar una experiencia. Hace algunos años, había un artículo que tenía un número muy considerable de citas, lo que me llamaba mucho la atención, incluso al propio autor le llamaba la atención, así que nos dimos a la tarea de buscar por qué se citaba tanto, para lo cual lo más fácil era encontrar los artículos que lo estaban citando, y nos llevamos

una sorpresa muy desagradable, porque descubrimos que la razón de tantas citas era que los otros autores decían «no hagan lo que este tonto hizo». Es decir, las citas eran malas referencias. Entonces, creo que habría que dimensionar muy bien las razones por las cuales los trabajos reciben citas. Dejando esta experiencia de lado, me gustaría puntualizar cuáles son las razones por las que nosotros pretendemos evaluar los artículos. Y ubico dos grandes razones: primeramente la evaluación que realizan los colegas, los pares, los que trabajan en la disciplina, ya que ellos van a tener criterios muy específicos sobre la valoración que hacen de los trabajos; y la segunda es la que nosotros, los administrativos, los bibliotecarios o la gente que trabaja en agencias de financiamiento o de gestión de la investigación podemos hacer. Me parece que las dos perspectivas son opuestas, porque mientras los que trabajan en la disciplina se ocupan de la pertinencia, del aporte, de la calidad científica, nosotros, los administrativos, nos ocupamos del número de descargas, del número de citas, del factor de impacto, etc. Entonces creo que esta posible oposición que se da entre las dos perspectivas, lo único que hace es que va ensanchando estas diferencias, porque algo que veo con mucha frecuencia es que los investigadores reclaman que no se hace una evaluación más detenida de sus trabajos, que los evaluadores se quedan en una visión muy superflua al tratar de establecer los distintos indicadores, y entonces se empiezan a acentuar algunos avances en ciertos parámetros que ya existen, no a surgir nuevos. Y entre ellos creo que podríamos valorar la importancia que tiene la honestidad académica, la honestidad científica, la pertinencia que tienen los trabajos, la honestidad que pueda haber dentro de las disciplinas y por supuesto, en ese sentido, el aporte que se hace a la disciplina es algo que estamos visualizando con mucha atención. Es lo que recientemente se acaba de reportar por parte de WoS, cuando nos dice que en América Latina los investigadores no están colaborando entre ellos: los investigadores mexicanos no saben lo que están haciendo los brasileños y los brasileños no saben lo que están haciendo los argentinos, cuando realmente todos tenemos temáticas en común, tenemos problemáticas en común, porque nuestra cultura es común y estamos prefiriendo colaborar con el Norte, con Europa y Oriente, y no estamos

volteando a ver lo que está pasando con nosotros. Entonces, quizá otra de las variables que tendríamos que estar midiendo es la colaboración regional y, por supuesto, lo que ustedes ya bien han comentado: la accesibilidad, la utilidad social y algo que también me parece importante, que es la internacionalización. En este sentido, la internacionalización la entiendo como la capacidad de hacer investigación al más alto nivel, en donde lo realmente importante es el aporte científico y no tanto la búsqueda de un reconocimiento en términos de financiamiento o de posicionamiento dentro de la academia.

Más allá de todos estos indicadores de los que estamos hablando, un problema que surge cuando uno considera nuevos indicadores es que se introducen nuevas dificultades (cómo se miden, de dónde se obtienen) y por lo general se empieza a ver que aparecen herramientas, servicios, librerías, software, etc.; entonces, quisiéramos saber si están al tanto de las dificultades a la hora de conocer estos nuevos indicadores para evaluar y si nos pueden mencionar algunas plataformas, herramientas, servicios, iniciativas, que ayuden a resolver estos problemas.

S. CAREGNATO: Simplemente son dos las mayores dificultades para considerar los nuevos indicadores: una es la falta de claridad sobre lo que un indicador está midiendo y de hecho si se consigue medir lo que él pretende; eso sucede mucho con las altmetrics, si se quisiera medir el impacto social con ellas, en realidad parece que medimos el impacto de ellas en los medios (redes) sociales, lo que es diferente. Y la segunda dificultad es la existencia de pocas herramientas que ofrecen metadatos de calidad y de formato abierto y gratuito para calcular esos indicadores. Mi preocupación es con las bases de datos y con las herramientas con las que se hicieron esos indicadores. Sobre la primera dificultad, la falta de claridad también existe en los indicadores tradicionales, los indicadores de citación son muchas veces utilizados como sinónimo de calidad y la calidad es un concepto mucho más amplio. Los indicadores de citación miden el uso, efecto e impacto en la literatura científica. Y con respecto a las altmetrics, hay todavía más dificultades para entender qué mide cada una de ellas, porque se basan en fenómenos recientes y poco estudiados, como mencioné ya, el impacto social de la ciencia o que nos gustaría medir,

porque lo que hemos medido con las altmetrics es el impacto en los medios sociales. ¿Por qué sucede esto? Porque las menciones de trabajos científicos en los medios sociales pueden suceder por diversas razones; muchas veces porque los títulos son divertidos, inusitados o peculiares; es decir, que son muchas las razones por las cuales se puede citar o realizar algún uso en los medios sociales. Esto se soluciona con más investigación y estudio, y con mucho debate sobre el resultado de esas investigaciones con los profesionales de la información, editores y comunidades científicas. Y sobre el segundo problema, sobre las pocas herramientas que ofrecen metadatos de calidad, abiertos y gratuitos para calcular los indicadores, en los últimos tiempos, he visto surgir varias nuevas fuentes de datos de citación, que son más inclusivas, más abiertas y con metadatos de calidad. Como mencioné, WoS y Scopus ofrecen metadatos de calidad, pero son bases muy limitadas. Ya las herramientas para calcular el impacto en los medios sociales, muchas de ellas son privadas, son servicios comerciales; por ejemplo altmetrics.com, PlumX o Mendeley. Pero felizmente nosotros tenemos un camino para solucionar eso y me parece una iniciativa muy importante en ese sentido, que es la Iniciativa por Citaciones Abiertas (Initiative for Open Citations, en inglés) que fue lanzada en 2017 y es un proyecto resultante de la colaboración entre directores, investigadores y otros interesados para promover la disponibilidad sin restricciones de los metadatos de citación y permitir su análisis. Los datos de citación son almacenados en la base de datos de CrossRef y CrossRef, además de hacer el registro de DOI, aloja también estos datos por lo que se transforma en un servicio colaborativo de indexación de archivos y de relación de las referencias entre esos archivos. Esto sin duda va a permitir que se tenga metadatos de calidad porque pasaron por la curaduría de los equipos editoriales de las revistas; va a estar disponible en un formato abierto y gratuito porque es el compromiso de CrossRef, que es una organización sin fines de lucro, y a la disponibilidad de los metadatos por vía API. A partir de aquí se pueden producir herramientas para crear varios indicadores de producción de colaboración, de citación y otros. Si bien no todas las revistas depositan todos los metadatos, o las referencias, pero nos estamos acercando a eso, nos

vamos acercando a una base de datos entrelazada que va a permitir ese uso. CrossRef ya permite el uso de sus metadatos. También existen otras fuentes como ScienceOpen, una plataforma abierta y gratuita, que utiliza datos de CrossRef, de Mendeley y otros proveedores, así como otros datos de lo que son los de producción y colaboración y citación. Lo que me parece claro es que estas métricas derivadas de los servicios comerciales no están alineadas con los objetivos de la ciencia abierta y no son suficientes para estas necesidades que tenemos hoy en día para evaluación de ciencia. En mi opinión, todo pasa por producir fuentes de información, herramientas con datos y metadatos detallados sobre los archivos, de calidad, abiertos y disponibles para uso por máquinas y por personas.

C. ROZENBLUM: Con respecto a la pregunta, coincido con Sônia y creo que vamos a coincidir todos en que una de las mayores dificultades es la falta de sistemas normalizados que incluyan todo tipo de metadatos y de documentos, pero sí creo que hay algunos sistemas que están trabajando con eso; de hecho, los mismos plugins de las revistas ofrecen muchísima información. El problema me parece que es unir esta la información y eso depende de para qué vamos a usar la información. La información está pero dispersa, está disponible en muchos lugares y en cada una de las de las distintas bases pero hay que saber buscar. Por ejemplo, una herramienta muy interesante es Publish or Perish, que es una herramienta abierta y que permite hacer búsquedas de citación bastante acertadas, pero no es considerada como un elemento que sea de gran utilidad para poder hacer la citación. Me parece que más que el problema de que no hay citas, hay que pensar qué quiero hacer con esta evaluación o qué quiero evaluar y entonces recoger de cada base de datos la información que necesito, porque, por ejemplo, una información de descargas de un artículo debería poder recogerla en el sitio de la revista, en Scielo, en Redalyc o en distintos lugares donde puedan juntar la información y eso va a depender del objetivo de la evaluación y de cada aspecto de evaluación que vaya haciendo. Entonces, simplemente eso: pensar que ahí hay muchísima información, que hay muchos sistemas que ya nos están ofreciendo mucha información normalizada y hay otros sistemas, otras herramientas que

nos permiten unir, además de muchas otras herramientas como las famosas API, que nos permiten ir bajando información y después cruzarla pero todo esto debería ser dependiendo del objetivo que yo tenga. Porque sí, está muy bien bajarme este dato, pero para qué me sirve si no lo puedo comparar. Si estoy evaluando investigadores, quiero ver cuántas descargas tuvo en el artículo de este investigador respecto del artículo de este otro o si quiero analizar revistas, etc. La información está, las herramientas están: lo que tenemos que pensar es que quizás el problema justamente es que estamos tratando de meter a la fuerza ciertos sistemas o ciertos indicadores en evaluaciones que no responden a ese objetivo. Sería bueno tratar de ir adaptándose a todo lo que sí tenemos de acuerdo a cada una de estas evaluaciones.

A. URIBE TIRADO: Bueno, para no decir como lo mismo, yo creo que hay dos aspectos claves: uno todo el movimiento de Open Citation es clave, porque en la medida en que no podamos capturar los datos para hacer el análisis, según lo que queramos o necesitemos, terminamos haciendo el análisis que podemos. Eso es lo que le ocurre a nuestras universidades y centros de ciencia y es porque siempre se utiliza WoS o Scopus, el primero porque está normalizado y lo que hacen relativamente lo hacen bien no se puede negar eso. Las críticas están en su cobertura y en sus costos pero el problema es que como es lo más fácil relativamente se lo sigue usando. Luego a Google Scholar, por ejemplo, hay que hacerle scraping para capturar los datos, es decir «robárselos». Es decir, la información está y está en múltiples fuentes, sean fuentes de la misma universidad como las descargas del repositorio, pero cabe preguntarse: ¿tenemos en los módulos de nuestros repositorios bien montado el asunto estadísticas y los editores o los autores lo pueden usar? Muchas veces no, entonces creo que las fuentes están pero hay que habilitar todo lo estadístico que éstas tienen y tratar de que todas las estadísticas pudieran estar abiertas. La misma crítica que se le hace a WoS o Scopus se le puede hacer a las cuatro empresas de altmetrics, les tenemos un mejor cariño porque miden cosas distintas y porque tienen algunas tienen API y entonces puedes capturar la información de ellos, pero hay otras que no tienen y si no se paga la suscripción, pues no se puede. La clave está en un movimiento abierto, tanto de

que capturemos de esos datos que ya existen, como abrir nuestros propios datos. Las revistas, por ejemplo, tienen sus datos estadísticos pero el módulo de OJS de estadísticas muchas veces no está habilitado o sólo se habilita el de cantidad de visitas, que no sirve para lo que nosotros queremos. Lo mismo ocurre con los repositorios, así que yo creo que hay que meterles el chip de lo abierto: abrir para consumir y capturar esos datos y poder hacer el análisis que queremos, pero también que lo que nosotros producimos lo abramos porque si no no sirve. Herramientas hay muchas: es decir, hace cuatro años hice un estudio y encontré que hay un documento que dice [«Las 101 herramientas de la ciencia hoy»](#) (está en internet) y hay una sección de métricas de evaluación y hay un montón uno puede utilizar y son buenas, pero es un fragmento de la fotografía. Yo utilizo WoS o Scopus, o utilizo Google, que es el fragmento mayor pero tiene problemas de indexación, y asimismo LENS u otras, el mismo DOI para empezar a hacer rastreo, pero creo que el reto está en integrar lo que tienen otros para capturarlo y abrirlo. Si no hacemos eso, nunca vamos a lograr la fotografía panorámica: saber qué es lo que queremos y saber qué queremos medir. A veces estamos midiendo cosas que no que no hay manera de medirlas, por ejemplo, cuál es el mejor profesor de una facultad: ¿es mejor el más citado por colegas o el que los estudiantes le entienden mejor o el profesor que en tesis de pregrado y posgrado más aparece o el profesor que hace más recursos para la universidad por proyectos de extensión? Es decir creemos que las citas son todo y sólo son un fragmento, porque las usamos porque es lo más fácil. Así que ahí está el reto: herramientas hay muchas, pero creo que hay que integrarlas y hacer propuestas, no seguir en la queja. Hacer nuevas herramientas sea desde nuestros contextos para tratar de compartir y tratar de integrar al máximo.

R. VILLEGAS TOVAR: Retomo un poco los puntos que pretendía abordar en mi intervención previa en cuanto a factores de valoración de los de los documentos y parto de cómo medimos la honestidad científica. Me parece que este es un concepto complejo en el que tenemos meras aproximaciones; en cuanto a herramientas digitales tenemos el caso de los sistemas de detección de plagio, que sabemos que se caracterizan por tener la capacidad de

identificar ciertos grados de similitud respecto a otros textos. Sin embargo los resultados que nos muestran ese tipo de herramientas invariablemente tienen que pasar por un proceso de análisis de interpretación por parte de especialistas y creo que esta es una situación que se repite constantemente dentro de los diferentes criterios de valoración: la importancia de la interpretación de esos resultados numéricos. Continuando con otro de los parámetros de valoración, podría cuestionar lo relacionado a la pertinencia: qué tan pertinente es un documento científico; dentro de esta pregunta me surge lo que conocimos en 2014 que hace *The British Medical Journal*: dentro de sus grupos de árbitros han abierto la puerta para que los propios pacientes sean quienes estén evaluando los trabajos de los investigadores, porque ellos son quienes, padeciendo estos malestares físicos, pueden evidenciar si esa investigación hace un aporte, hace un avance o simple y sencillamente está repitiendo más de lo que ya se ha dicho. Otro cuestión: cómo medimos la colaboración científica, cómo podemos nosotros dimensionar los impactos de la diplomacia científica y creo que la respuesta sería simple: contabilizando el número de acuerdos de colaboración que existen entre las universidades; sin embargo cuando yo he hecho este ejercicio ha sido infructuoso porque me doy cuenta que comúnmente los investigadores colaboran de manera científica sin hacer acuerdos de colaboración, trabajan más como un ejercicio de investigación y no tanto como consecuencia de una serie de protocolos institucionales. Tan es así que nosotros en las instituciones nos enteramos de las colaboraciones una vez que se ha publicado el trabajo y no cuando se está haciendo. Otra variable podría ser la accesibilidad y de nueva cuenta retomo un caso que tuve hace poco tiempo en donde un alumno que no conocía muy bien el manejo de fuentes de información me decía que se había enfocado exclusivamente en documentar su trabajo de tesis en trabajos que estuvieran en acceso abierto, entonces pareciera ser que en esta apuesta que hacemos al acceso abierto estamos dejando por atrás lo que pudiera estar en acceso por suscripción, cuando es igualmente importante. Otra variable tendría que ver con la captación de recursos: hoy inclusive se miden las instituciones por su capacidad de captación de recursos económicos para financiar la

investigación, ejemplo de esto es lo que sucede con el ranking Times Higher Education, que dentro de sus variables miden precisamente la capacidad de captación de financiamiento por parte de sus investigadores, y la pregunta es: ¿vamos a medir la investigación en función de la capacidad económica que tengamos de captar recursos? Eso me parece bastante complejo porque entonces instituciones que no tengan esta capacidad de captación como las puedan tener en el Norte prácticamente no tendríamos nada que hacer en la escena científica. Entonces, insisto, y con esto concluyo: me parece que el reto está, por una parte, en tener los criterios y, por otra, en interpretar los resultados numéricos que nos arrojan.

Aprovechando que hay participantes de varios países y que estamos hablando tanto de todos estos cambios que están surgiendo (nuevas formas de evaluar, nuevas métricas), nos interesa conocer si saben si se está realizando algún esfuerzo de proyectos o leyes para cambiar (bien en sus países o en sus instituciones y ámbitos de trabajo) los métodos de evaluación científica.

S. CAREGNATO: Es un asunto muy difícil de comentar en Brasil, de momento, porque estamos en el proceso de evaluación cuatrienal: cada cuatro años la agencia responsable del sistema de graduación hace el ejercicio de evaluación. Algunas cosas han cambiado en las últimas décadas principalmente por causa del tamaño de la producción brasilera, por el crecimiento y por la cantidad de producción científica de los investigadores. Nuestro sistema de evaluación de graduación es complejo y lo mismo su evaluación cualitativa y cuantitativa. Estos cambios no fueron suficientes para incorporar en las evaluaciones los nuevos tipos de publicación y los nuevos indicadores. Lo que cambió recientemente en Brasil y está siendo discutido es Qualis, un instrumento que da un ranking de revistas, en el que los investigadores brasileños publican y además del ranking da una estratificación para ellos. En 2019 se propuso una nueva metodología derivada de Qualis, a partir de criterios más objetivos. Anteriormente, tenía una evaluación cualitativa de las revistas y ahora usa exclusivamente los criterios cuantitativos, que se basan en cuatro principios: una clasificación única, una clasificación por grandes áreas del conocimiento,

referencias e indicadores bibliométricos. Ahora se está proponiendo que se usen tres tipos de indicadores: CiteScholar de Scopus, el factor de impacto de WoS, y el índice a5 medido por Google Scholar, es decir, todos son indicadores de citación. A pesar de los grandes esfuerzos para mejorar Qualis, persisten los problemas conceptuales, en la metodología y en la toma de decisiones. Las revistas tomadas aisladamente no son suficientes para medir la producción científica. Qualis ha recibido críticas por unir esos tres indicadores distintos: por un lado, el índice a5 parece beneficiar a las revistas de ciencias sociales que no están indexadas en WoS o en Scopus, por otro lado no siempre los resultados de Google Scholar son confiables; en este sentido, son beneficiadas revistas que tienen un mayor *size*, es decir, que publican un mayor volumen, más trabajos durante el año. Es un momento de mucha controversia y la gente está viendo que las formas de medición tradicionales al menos nos dan métricas cuantitativas.

C. ROZEMBLUM: Nosotros en la Argentina tenemos la buena y la mala suerte de no tener un sistema explícito de evaluación. Lo más próximo es la evaluación que se hace en CONICET, pero tampoco tiene unos lineamientos claros, precisos y explícitos. Como decía esto puede ser una ventaja, porque eso nos da la posibilidad de pensar que podemos hacer en algún momento una propuesta muy inclusiva, pero también nos deja un poco en el aire. Lo que sí me parece que es interesante resaltar es una resolución que hubo en el 2014, aprobada por el directorio de CONICET que planteaba la jerarquización de las revistas a partir de su inclusión en bases de datos, al menos para las de Ciencias Sociales. La resolución es un documento es un documento muy interesante pero, nuevamente, era una sugerencia a las comisiones. En esa categorización se trabajó mucho y se logró (fue un trabajo muy arduo y muy discutido) que por ejemplo en el nivel 1, donde se guiarían solamente por WoS y Scopus se logró poner a SciELO. Personalmente, yo creo que en Latinoamérica una revista que está en SciELO o en Redalyc merece todo el respeto de toda la comunidad porque ha pasado por miles de evaluaciones. Actualmente, hay una comisión que está trabajando en ciencia abierta para el Ministerio de Ciencia y Tecnología, y está trabajando sobre algunos lineamientos y sugerencias en ese

sentido. Esperemos que salga de ahí algo que pueda activar algún tipo de análisis y de resolución más explícita para las evaluaciones. Por ejemplo se está trabajando en sugerir enfáticamente a los investigadores nacionales que publiquen un artículo o al menos una contribución en revistas locales o regionales como una forma de incentivar justamente esta diversidad de canales de información. Esto es lo que tenemos efectivo en Argentina.

A. URIBE TIRADO: Creo que hay un primer reto: convencer a nuestros propios directivos de nuestras propias universidades, alfabetizarlos en todos estos temas, porque muchas veces nuestros directivos o no conocen del tema o prefieren quedarse en lo tradicional. Creo que ese que ha sido el logro en la Universidad de Antioquia (Medellín, Colombia), porque hemos logrado un grupo de investigadores y grupos de investigación que conocen nuestros temas generando cultura en esas directivas y en la propia institución de que podemos medir cosas distintas y hacerlo distinto. En el chat hemos puesto un enlace con la [tabla de homologación](#) para que todas las disciplinas puedan competir en mismas condiciones en nuestra universidad. El ejemplo obviamente es que escribir un artículo para las ciencias básicas es más fácil, no por lo académico sino porque hay más revistas, es un trabajo colaborativo, es de alcance mundial, se publica en inglés, etc. En cambio, es más difícil para las humanidades porque no siempre tienen que ir en inglés, no es lo pertinente siempre, no hay tantas revistas, etc., entonces ¿cómo se logra evitar esa queja? Pues tratando de tener herramientas que igualen las disciplinas, que lo hagan más equitativo. Esa tabla de homologación fue una discusión con los investigadores en la universidad, de decir no sólo pueden ser unas áreas las que les vaya bien, las que siempre se ganan la financiación, porque si siempre ponemos investigadores para artículos puros, siempre van a ganar las mismas disciplinas y no estamos en igualdad de condiciones, o hay investigadores que trabajan con investigadores de centros internacionales, sí, pero hay algunos que trabajamos más para América Latina por pertinencia y otros trabajan con el primer mundo porque es lo pertinente. Esas tablas de equivalencia tratan de generar equidad disciplinar en la misma institución. El otro enlace es algo que hemos llamado «Colav» y es empezar a hacer estudios de vinculación, ya que

queremos medir cosas distintas y no queremos quedarnos sólo en la citación: pues empezamos a medir otras cosas y crear herramientas que sirvan para esto. Son cosas que todavía están en prueba, pero que demuestran que sí se pueden medir cosas distintas, es un trabajo colaborativo para medir otras cosas: buscar respuestas como con qué empresas trabaja la universidad y qué le ha aportado y qué bancos o qué ONG se sirven de las investigaciones de nuestras instituciones, y **allí** ya se pueden encontrar. Finalmente, terminaría con un esfuerzo, que es el **FOLEC** con CLACSO: América latina tiene que medir su ciencia con una mirada global, pero no olvidar lo local y regional. Nuestro error es ese: nuestro sistema de ciencia quiere medirla como se mide en Europa y no puede ser, pues es injusto epistemológicamente y sociogeográficamente.

R. VILLEGAR TOVAR: Pues brevemente quiero comentar que en México hemos estado transitando esto de medir la productividad y el impacto y ahora nos estamos volcando a dimensionar la importancia de la ciencia abierta y aspiramos a movernos al impacto social. Claro que la gran inquietud o la gran duda es cómo vamos a medir el impacto social. Pues estamos en esa etapa y creo que, como todos, estamos también en este proceso de aprendizaje.

Mesa 3: Datos de investigación en abierto

IANKO LÓPEZ ORTIZ DE ARTIÑANO

Director técnico, Consorcio Madroño (España)

HUMBERTO BLANCO CASTILLO

Universidad del Rosario (Colombia)

RAÚL SIFUENTES ARROYO

Unidad de Automatización del Sistema de Bibliotecas, Pontificia Universidad Católica del Perú (Perú)

RAFAEL PORT DA ROCHA

Proyecto Red de Datos de Investigación Brasileña (RDP), Universidade Federal do Rio Grande do Sul (Brasil)

Modera: **MALGORZATA LISOWSKA NAVARRO**

Universidad del Rosario (Colombia)

RESUMEN EXTENDIDO

IANKO LÓPEZ ORTIZ DE ARTIÑANO (España): Hay una cuestión que relevante cuando hablamos del Consorcio Madroño y es que en su ADN está incluida la ciencia abierta; de hecho en el plan estratégico actual se establece como misión básica principal del Consorcio proporcionar una infraestructura de información que impulse la excelencia de la investigación en las instituciones miembro, contribuyendo al desarrollo de la ciencia abierta, a la transformación digital y al desarrollo sostenible. Mejorando la experiencia de sus usuarios. Es decir que la propia misión del Consorcio ya incluye la ciencia abierta y asimismo la visión, porque lo que queremos es ser una organización líder y, entre otras cosas, buscamos la promoción de la ciencia abierta, la transformación

digital y el desarrollo sostenible. El Consorcio se creó en 1999, con siete miembros originales, todos con los objetivos de incrementar la producción científica de sus universidades, mejorar la calidad de los servicios y ahorrar costes, promover planes de cooperación entre sus miembros y la aplicación de nuevas tecnologías. Por lo que se refiere específicamente al primero acceso abierto, después a la ciencia abierta, se ha seguido un determinado determinado camino, que comienza en 2005 con la definición de e-ciencia, que es una red de repositorios institucionales de acceso abierto de tal manera que cada universidad miembro del consorcio y también otras instituciones mantienen sus repositorios de acceso de acceso abierto de contenidos pero el Consorcio Madroño los aglutina en una red de repositorios común, apoyada por la Comunidad de Madrid, tanto financiera como operativamente, que es [e-ciencia](#). A lo largo del tiempo en estos últimos 15 años pues han ido sucediendo varios pasos, entrar en cada uno de ellos posiblemente sea una tarea muy exhaustiva, pero es especialmente relevante llegar a 2016, cuando se crea [e-cienciaDatos](#), que es un repositorio institucional para almacenar datos y poner en abierto datos de investigación, un proyecto pionero en el ámbito español. Inmediatamente después, todos los servicios relativos a ciencia abierta se integran en un mismo portal llamado [investigaM](#), que a su vez está inserto en la web del Consorcio Madroño. Este portal incluye tanto e-ciencia, como e-cienciaDatos, como la herramienta PaGoDa de PDG on line que permite a los investigadores la realización de planes de gestión de datos según las premisas establecidas por Horizonte 2020 de la Comunidad Europea, ahora Horizonte Europa. También en 2017 el Consorcio Madroño emite una declaración de apoyo la ciencia abierta con recomendaciones para los investigadores, para las universidades y también para los gobiernos y las instituciones públicas donde, y esto es importante, se mencionan ya expresamente los datos de investigación. En la actualidad hay aproximadamente unos 570.000 registros en e-ciencia, en e-cienciaDatos unos 500 a 600 datasets, 850 PDG y seguimos creciendo en todos estos aspectos. Como comentaba anteriormente, todos estos servicios y toda la infraestructura de ciencia abierta del Consorcio Madroño, todos estos desarrollos se integran

en el portal investigaM, que es el portal de ciencia abierta. Específicamente, e-cienciaDatos se estructura como un sistema que está constituido por varias comunidades que agrupan los datasets de una de las universidades miembro. E-cienciaDatos ofrece por tanto el depósito y también la publicación de conjuntos de datos asignándole un identificador de objeto digital (DOI) a cada uno de estos datasets. La asociación de un dataset a un DOI facilita la verificación de los datos, la diseminación, su reutilización, el impacto y el acceso a largo plazo. Además el repositorio provee una cita normalizada para cada dataset, lo que contiene información suficiente para que éste pueda ser identificado y localizado en un momento dado incluyendo el DOI. Los datos, por tanto, están en acceso abierto. Como decíamos, se trata de un repositorio de datos multidisciplinar, que alberga conjuntos de datos científicos de los investigadores de las universidades miembro del Consorcio Madroño, para darles visibilidad, garantizar su preservación y facilitar su acceso y reutilización. Se basa en el sistema Dataverse, con una comunidad/dataverse por cada universidad, gestionada por la propia universidad, y comunidades internas por proyectos. También tiene un gestor de estadísticas. Además de esto, cuenta con el sistema de PDG on line que, como decía, es una herramienta que permite a los investigadores realizar, de manera tutorada, planes de gestión de datos según las premisas establecidas por Horizonte 2020, hoy Horizonte Europa.

JUAN CORALES (España): Voy a presentar el repositorio de datos e-cienciaDatos, lo haré muy rápidamente porque vamos justos de tiempo. E-cienciaDatos es un repositorio de datos muy útil y multidisciplinar. Pueden subir sus datos los investigadores de todas las universidades del Consorcio Madroño. Tenemos una comunidad por cada una de las universidades que forman parte del Consorcio y está dirigido a las necesidades de nuestros investigadores; principalmente, ofrece servicios como geolocalización, ya que la mayoría de los datasets están localizados, ofrecen por supuesto una dirección de correo de soporte, un manual, y otra característica que nos han pedido siempre los investigadores es que sus datos tienen que ser visibles. Por esta razón es recolectado por los principales harvesters de datos, como el

propio Dataverse de Harvard y otros; también, otra funcionalidad que necesitan nuestros investigadores, es que sea un repositorio válido tanto para las principales editoriales como para las agencias de financiación. Entonces nos hemos asegurado de que e-cienciaDatos pueda ser utilizado para almacenar los datasets de investigación; es decir, si alguna revista pide que estén en un repositorio temático. Si vemos ya un dataset en concreto, pues también damos otras herramientas interesantes para los investigadores, como estadísticas del dataset, por supuesto, lo que ha nombrado lanko del DOI, que es indispensable, una cita estándar que se puede exportar en varios formatos y un potente gestor de versiones. Las estadísticas, aparte de ser importantes para los investigadores, también son importantes para nosotros como gestores de datos; también tenemos acceso a ellas en dos formas: una es la propia de Dataverse que nos da información sobre todo de cómo se usa e-cienciaDatos, a nivel de cómo ha ido evolucionando el uso de ficheros de datos, de las principales áreas temáticas y también tenemos unas estadísticas de Matomo, que nos da muchísima información sobre desde donde se accede, sistemas operativos, navegadores, muchísima más información de la que podemos navegar.

RAFAEL PORT DA ROCHA (Brasil): Nosotros pertenecemos al Grupo de Trabajo Rede de Dados de Pesquisa (GT-RDP) del Centro de Documentação e Acervo Digital da Pesquisa (CEDAP), que tiene como objetivo la digitalización y repositorios de datos de investigación. Tenemos un enfoque, un framework conceptual en el que se trabaja con el objetivo y la perspectiva de ser un repositorio digital confiable. Nos preguntamos inicialmente: ¿dónde está la información, los criterios, los requisitos que se tienen que identificar para planear y desarrollar el repositorio? Encontramos el modelo Open Archives y un mecanismo de certificación al que deberíamos someter el repositorio (FAIR, ACTDR ISO 16363 e ISO 14721). Nos propusimos entonces montar el proceso de desarrollo del repositorio para que todo lo que no se había conseguido lograr hasta entonces se comenzara a realizar con la idea de obtener un repositorio digital confiable. Pensamos los pasos de implementación del repositorio

teniendo en cuenta los requisitos de repositorios digitales confiables. Un repositorio no es solamente un software y las normas de repositorios son muy importantes para la gestión de riesgo. ¿Dónde están los riesgos que hacen peligrar la continuidad de un repositorio? Los riesgos son tecnológicos pero también organizacionales. La investigación que llevamos a cabo abarcó DOI, Data Cite, Dataverse, infraestructura, archivemática y preservación digital, entre otros aspectos. Investigamos también la comunidad académica productora de los datos. Trabajamos mucho respecto a la comunidad productora de datos y encontramos que muchos investigadores desconocen el contexto de repositorios y repositorios de datos. Estudiamos Dataverse y DSpace y los trabajos de ingeniería de software y criterios de evaluación de software con el objeto de tener un cierto dominio, seguridad y argumentos para el uso de esas herramientas enfocadas en repositorios de datos. Los repositorios de datos son algo nuevo con respecto a los repositorios de documentos, que se conoce bien cómo funcionan, cómo se producen y se consumen los documentos, las lógicas de los artículos, las tesis, las tesinas y cómo se produce ese conocimiento. Con respecto a los repositorios de datos buscamos conocer cómo se produce ese conocimiento, esos datos, para no crear un repositorio que no sea usado. Como resultado de los primeros estudios y en conjunto con la Red Nacional de Investigación (Rede Nacional de Pesquisa) y con el IBICT en el lanzamiento de un proyecto de incubación de repositorios. Estamos apoyando a cuatro instituciones para el desarrollo de cuatro repositorios con esta perspectiva. Se colocaron ciertos requisitos, responsabilidad, autodirección, equipo de ciencias de la información y equipos de investigación.

HUMBERTO BLANCO CASTILLO (Colombia): La experiencia de la Universidad del Rosario arrancó con un diagnóstico de un grupo de investigadores, con el fin de conocer la forma en que ellos estaban generando sus datos, dónde los estaban almacenando y qué hábitos tenían sobre los datos, sobre su respaldo y demás. A partir de esto realizamos una encuesta y luego fuimos realizando diferentes ejercicios para obtener esta información, validar algunos supuestos iniciales y entender ese comportamiento que tenían los investigadores para poder

incluirlo en nuestra estrategia de gestión de datos. De aquí salió una trivía que llamamos «Datos de investigación: ¿ya eres un pionero?», una encuesta que llamamos «¡Es el momento de tus datos de investigación!». También hicimos un *focus group* con los investigadores que detectamos que ya estaban inmersos en algún proceso que requería de datos de investigación. Ya con esta información, lo que encontramos dentro de los hallazgos fue que un porcentaje alto de los profesores no incorporaba los planes de gestión de datos dentro de sus proyectos, que los datos estaban almacenados generalmente en su computador personal o en una memoria USB sin ningún respaldo o los tenían en la nube, muchos de sus buzones personales. Respecto a compartir los datos los compartían únicamente con sus redes de colaboración y tenían algunas inquietudes respecto a reusar datos de otros. Con esta información más o menos organizada se inició el despliegue de la estrategia de gestión de datos y el primer paso consistió en la creación de un marco normativo a través del cual se pudieran establecer lineamientos para preservar y gestionar estos datos. Así es como en 2019 creamos la política institucional para la gestión de datos de investigación, con el cual se promovía esta apertura de los datos, acogiéndonos al principio de «tan abierto como sea posible, tan cerrado como sea necesario». Dentro del documento se reconoce la importancia de los datos de investigación, se hace la definición de algunos principios, y se definen los compromisos de la institución y los compromisos de los investigadores. De parte la institución, a brindar el acompañamiento y proveer los mecanismos para el almacenamiento y la identificación de los datos, y por parte de los investigadores, proveer datos de calidad. El siguiente paso fue crear el repositorio de datos y los servicios asociados. Respecto a la implementación del repositorio, se realizó directamente con el equipo técnico que tenemos en el CRAI; el software que utilizamos, al igual que en las dos presentaciones anteriores en esta mesa, utilizamos Dataverse, porque se adaptaba mucho mejor a los requerimientos que habíamos planteado, en términos de los aspectos básicos de almacenamiento, explorar los datos, enlazarlos, citarlos, el tema del versionamiento, que también ya se mencionó, y además estábamos buscando una plataforma que fuera de código abierto, que contara con una

comunidad robusta y que permitiera, digamos, «tocar» el código de una manera más fácil para el desarrollo de plugins y que contara con buenos estándares de interoperabilidad; por estas razones seleccionamos Dataverse. Luego, una vez que seleccionamos el software que íbamos a utilizar, debíamos definir la estructura a través de la cual íbamos a almacenar los datasets dentro del repositorio y, luego de explorar varias opciones, la que mejor se ajustaba a la forma como íbamos a desplegar el servicio era a través de proyectos: es decir cada dataverse corresponde a un proyecto de investigación y dentro de él se colocan cada uno de los datasets. También era importante definir cuál iba a ser el modelo de administración y para esto se determinó que el investigador principal, previo una capacitación que se le entrega, respecto a la administración interna del repositorio, iba a ser el administrador del dataverse, pues él ya iba a generar los permisos para cada uno de los investigadores o colaboradores para que subieran su información. Esto porque dependiendo de cada proyecto, la forma de gestionar los datos va a ser diferente o bien los proyectos pueden estar conformado por diferentes equipos, por diferentes grupos y por lo tanto pues variaría ese modelo de de permisos. En noviembre de 2019 pudimos poner a disposición de la comunidad el primer dataverse en abierto, llamada «Leishmania in the Américas DB». Dentro de este proceso de implementación, también se realizó la implementación de algunos plugins, como los de visualización de estadísticas, la visualización de hojas de datos tabulares y, en este caso, por ejemplo, los servicios de datos de geolocalización. Cuando hicimos la migración a la versión 5 tuvimos algunos temas respecto al despliegue, porque Dataverse cambió el modelo de visualización de estos datos. Otro aspecto importante fue la adquisición del DOI para la identificación de los datasets; en este caso nos suscribimos a Dataverse como miembros y esto permite, una vez que se configura, que los DOI se asignen directamente sobre los datasets una vez que se publican y se registra directamente ante DataCite y, además, esta información queda indexado directamente en Google Datasets, entonces esto nos pareció una de las cosas muy ventajosas para mostrar a nuestros investigadores. Respecto a los servicios, diseñamos e implementamos servicios que nos permiten soportar

algunas fases del ciclo de vida de los datos; por ejemplo, la primera fase, que es la fase de planeación, estábamos apoyando a los investigadores con la construcción del plan de gestión de datos y realizamos un acompañamiento a los profesores para la construcción de este documento; respecto a la etapa de almacenamiento y difusión ya se implementó un servicio de capacitación y asesoría, como les mencioné anteriormente, primero para uso del repositorio pero también hay aspectos de sensibilización acerca de la importancia de la apertura de los datos, el uso de los datos de investigación y la publicación de los datos.

RAÚL SIFUENTES ARROYO (Perú): Voy a compartir la experiencia de la Pontificia Universidad Católica del Perú en datos abiertos. Menos mal que los colegas que me han antecedido ya presentaron algunos conceptos relacionados a esto, porque como corre el tiempo, he sacado algunas cosas que ya se han tocado. Vamos a ver por qué escogimos Dataverse, cuáles son las características actuales que tiene el repositorio y cuál es la relación que tiene con el repositorio institucional, qué proyectos están en marcha y cuáles son los futuros pasos. y Nosotros escogimos Dataverse, en realidad, por una experiencia previa que antecede más o menos al año 2013, cuando lanzamos el repositorio institucional de la Pontificia Universidad Católica del Perú. Hicimos también una instalación de Dataverse porque en ese momento un instituto de investigación buscó el apoyo de la biblioteca para poder hacer la implementación y ellos conocían muy bien Dataverse, tenían experiencia ya por haber utilizado esto en Harvard. Sin embargo, después del lanzamiento del repositorio el Dataverse que que tuvimos se tuvo que dejar de lado, porque la persona que movió todo este tema salió de la universidad y se fue a otro trabajo, entonces nosotros nos quedamos con el tema de Dataverse por allí, digamos un poquito con la miel en la boca pero ya lo habíamos experimentado, ya habíamos hecho la instalación y todo. Pasaron seis años para que otro instituto de investigación nos busque y nosotros le ofrecimos, como una solución, el Dataverse para lo que ellos estaban buscando. Entonces en 2019 hicimos el lanzamiento del Dataverse porque necesitábamos tener esa

experiencia. ¿Por qué nos metimos en ese tema? Porque la legislación peruana así como ya exige tener en las instituciones repositorios de tesis o repositorios institucionales, también en algún momento ya va a empezar a exigir que las instituciones como las universidades tengan repositorios de datos, entonces queríamos tener esa experiencia, tuvimos el aliado pertinente para ese momento y lo lanzamos en septiembre del 2019. después de haber analizado todos los programas que hay en el mercado escogimos Dataverse porque para nosotros tenía los metadatos muchos más desarrollados en las áreas que nosotros tenemos en la universidad, que son las ciencias sociales y humanidades, que es muy fuerte en nuestra universidad, astronomía, astrofísica, ciencias de la vida y espaciales, y además revistas. Es decir, cumplía con nuestros requerimientos y además vimos que había una creciente comunidad en español. Me alegra ver que ahora hay una nueva instalación en Ecuador; en Perú tenemos dos oficialmente, la de la Universidad Católica y un Centro Internacional de la Papa en Colombia, hay otra en Chile; vemos que en Brasil hay varias instalaciones, hay en México también y en el Caribe una universidad de las Antillas también. Una de las cosas por las que nos elegimos Dataverse también fue que se integraba con el OJS. ¿Qué características actuales tiene? Ya mencioné que fue lanzado en septiembre del 2019, tenemos actualmente 6 repositorios, 92 conjuntos de datos, la mayor parte de estos conjuntos de datos son entrevistas realizadas por este Instituto de Opinión Pública que a lo largo de su historia ha venido haciendo encuestas electorales y encuestas sobre temas importantes para la coyuntura del país, algunas veces encargadas por terceros y otras veces por la misma motivación de la universidad. Actualmente tenemos 541 archivos y utilizamos como ID persistente el handle; estamos utilizando Amazon Web Services para el software alojado ahí y algo que les quiero comentar es que tanto el repositorio de datos como el repositorio institucional, las revistas y todo todo lo que tenga que ver en realidad con la ciencia abierta está gestionado bajo el Sistema de Bibliotecas de la universidad que tiene a su cargo los servidores y además el servicio profesional para brindar los servicios desde el punto de vista de las bibliotecas. Entonces tenemos ahí un área de informática especializada, que es

el área que yo dirijo para todos estos temas. Otra de las características actuales es que estamos indizados en Google Dataset Search, como mencionó mi colega de la Universidad del Rosario, ellos utilizan el Data Cite y eso les permite entrar directamente en Dataset Search. Nosotros tenemos que actualizar, creo que es una primera vez se tiene que ir mandar un archivo XML a Google Dataset para la indexación. Hasta el día de ayer hemos registrado 5.598 descargas de los seis repositorios que mencioné. Hay una relación con el repositorio institucional, pues todos los datasets están cosechados por el RI porque el RI es la fuente principal de metadatos para el nodo país que vendría a ser el CONCYTEC, que es un aportante de LA Referencia. Esos metadatos se cosechan del repositorio institucional y se enriquecen según la legislación vigente nacional. Tenemos una legislación que nos exige la presencia de ciertos metadatos. ¿Cuáles son los proyectos en marcha y a futuro? Ahora mismo tenemos Mapas del Perú: se están creando bases de datos de mapas del siglo XIX o XX que les sirven a varios investigadores para poder estudiar justamente cómo han cambiado los mapas; hay mapas mineros, mapas de petróleo, mapas de carreteras, de caminos, etc. Incluso recuerdo que hace poco nos pidieron unos mapas para resolver un conflicto de tierras que se había originado en una zona del norte del país. Tenemos el Archivo Digital de Lenguas Peruanas que son estudios, audios, videos, transcripciones de documentos fonéticos, diccionarios, y es el único archivo digital de las lenguas nativas del Perú. Ha sido llevado a cabo por los lingüistas de la universidad y también otras investigadores. Tenemos bases de datos de poetas y eso más o menos representaría 50 datasets y 800 archivos. Para terminar, estamos elaborando los archivos bibliográficos en formato RIS de todas las revistas para ponerlos on line y para que generen interés en investigación bibliométrica de todas nuestras revistas. También estamos pensando poner datos transaccionales estadísticos y de gestión de las bibliotecas porque nos piden mucho las personas que quieren hacer estudios de usuarios, y también vamos a integrar las revistas de la universidad. Además algo que está muy cerca es que se va a incluir los datos en una futura política de ciencia abierta en la que vamos a mirar la experiencia de la Universidad del Rosario que nos presentó este año en

algún evento hace unos pocos meses. Como ustedes pueden ver, tenemos datos de la investigación y también para generar investigación; también datos crudos que buscamos que generen investigación como en el caso, por ejemplo, de los registros bibliográficos de las revistas.

Mesa 4: Infraestructuras tecnológicas para gestión de datos de investigación

LUIS ELIÉCER CADENAS MARÍN

Director Ejecutivo de Red CLARA (Venezuela)

MANUELA KLANOVICZ FERREIRA

Analista de Sistemas del Centro de Procesamiento de Datos, Universidade Federal do Rio Grande do Sul (Brasil)

LAUTARO MATAS

Secretario Ejecutivo y Técnico LA Referencia (Argentina)

Moderadora: **MARISA R. DE GIUSTI**

Universidad Nacional de La Plata (Argentina)

RESUMEN EXTENDIDO

¿Qué infraestructuras de datos y a qué niveles considera importante desarrollar en un país: regional, nacional y/o institucional? ¿Por qué? ¿A qué nivel se deben manejar los grandes volúmenes de datos generados?

LUIS ELIECER CADENAS (Venezuela): Es una pregunta compleja, diría yo, que tiene más de una respuesta, y depende eventualmente de muchos factores y también del nivel, tanto geográfico como de alcance que se piensa, como las disciplinas sobre las cuales estamos tratando de manejar o utilizar estos datos. Desde Red CLARA he tenido la oportunidad de interactuar con múltiples iniciativas globales que están trabajando el tema de los datos científicos, en el marco de lo que se ha denominado «nubes de ciencia abierta» y que hay ya varias experiencias y esfuerzos a nivel global. Eso incluye, por ejemplo, la experiencia europea, la European Open Science Cloud, incluye el proyecto africano de Open Science Cloud, y múltiples iniciativas que buscan generar

consensos en términos de cómo manejar los datos de una manera que puedan hacerse interoperables y podamos sacar el mayor provecho en términos de inversión que hacemos en la región. El elemento importante a considerar es lo que representan esos datos en términos de inversión por parte de gobiernos en toda nuestra región y la relevancia que tiene les saquemos el mayor provecho posible. Yo pienso que en término de los niveles, debe haber infraestructura a todos los niveles, debe haber infraestructura local, dependiendo y justamente por las características de desarrollo tecnológico o de digitalización que tiene Latinoamérica y la gran variedad de niveles, conocimientos y capacidades que tienen las propias universidades, tenemos que ayudar a que esto ocurra de una manera más o menos generalizada y eso supone crear mecanismos de apoyo y soporte a esas instituciones para puedan fomentar y desarrollar su infraestructura de datos. Las grandes universidades las tienen, obviamente, y tienen resuelto este problema, pero queda una enorme cantidad de instituciones que eventualmente no conocen, no saben o no tienen idea de cómo mantener una infraestructura de esta naturaleza. Creo que hay que pensar en lo local, en donde lo más apropiado entiendo que es una integración de los datos ascendente, es decir, de lo más bajo hacia lo más alto en términos de integración; buscar mecanismos de interoperabilidad y coordinación para poder construir una visión de esos datos en una manera distribuida a nivel regional y yo pensaría, a lo menos, en una instancia que permitiera integrar esos datos a nivel latinoamericano. Construir los mecanismos para ayudar a esas instituciones pequeñas o que no conocen de este tema para que puedan fácilmente comenzar a almacenar y manejar sus datos de forma inteligente. En las instituciones grandes, dependiendo de las disciplinas hay mucha diferencia, o sea, no es lo mismo los datos que manejan en un repositorios los físicos de altas energías, como el CERN, que los datos que puede necesitar alguna de las ciencias sociales, ni la manera en que eso se almacena y se utiliza. Por eso decía que la pregunta es muy compleja, porque es un espacio muy fragmentado en todos los sentidos: en el sentido de las estrategias de las organizaciones están buscando solventar esta problemática o apoyar en mejorar el manejo de los datos, y después hay un componente que tiene que ver con estas

capacidades técnicas y humanas en cada una de las instituciones. Hay un componente que tiene que ver con las características de la ciencia sobre la cual o de la cual se están almacenando los datos. Es decir, es muy muy complejo y yo creo que hay que ser flexible en términos del modelo que se implemente y tener una especie de caja de herramientas con muchas alternativas para poder atender cada necesidad en el caso, en el contexto, con las condiciones y con las características que haga falta. Tratar de tener esta caja de herramientas lo más completa posible; seguramente Lautaro comentará luego alguna de las cosas que estamos haciendo conjuntamente la Red CLARA, LA Referencia, el Sistema de Cómputo Avanzado para América Latina y el Caribe (SCALAC) para crear un ecosistema que efectivamente apoye este manejo de los datos de una manera eficiente. Para concluir mi respuesta, quisiera comentar solamente la importancia de que miremos también las tecnologías emergentes de transformación digital y el impacto que puede tener el uso de tecnologías como la inteligencia artificial o las tecnologías de block-chain en el proceso de investigación y en el mejor aprovechamiento de los datos científicos. Hay una revolución alrededor de la capacidad para producir ciencia, gracias a la capacidad para gestionar y procesar los datos, y creo que esos componentes debemos tomarlos en consideración porque representan un cambio o pueden impactar de manera muy importante la competitividad o la capacidad productiva de nuestros investigadores en la región. Estamos ante un gran reto. Yo creo que están confluyendo muchos intereses, muchos actores, hay mucha cooperación internacional buscando fortalecer los sistemas científicos; hemos vivido en el marco de una pandemia que ha empujado una concepción y un entendimiento mucho mayor de lo que significa la ciencia y yo creo que es momento de ponernos todos de acuerdo y empezar a construir colectivamente, con flexibilidad, sabiendo que no vamos a tener una solución que resuelva todo y que vamos a tener que tener alternativas que puedan ayudar en los distintos tipos de problemáticas. Entonces creo que tiene que ser local, a nivel de la institución, nacional para integrar lo que tiene que ver con las instituciones, y tiene que ser latinoamericano y probablemente global, si queremos estar inmersos en lo que es la ciencia global y sus potencialidades y sus logros.

MANUELA KLANOVICZ FERREIRA (Brasil): Conuerdo con algunas de cosas que dijo Luis. Aquí, en Brasil, todas las infraestructuras, tanto locales como regionales como institucionales, son las tres importantes. Los investigadores tienen varias formas de trabajo, dependiendo del área de investigación, y muchas veces dentro de la misma área proyectos con recursos distintos trabajan de forma diferente. Entonces los tres niveles son importantes. Por ejemplo, en la infraestructura local o institucional, hemos tenido oportunidad de promover la preservación durante la investigación; es decir que los datos antes de ser publicados ya pueden ser preservados por los investigadores y esos datos ya tienen una curaduría, una procedencia documentada, de forma que cuando llegue el fin de la investigación y se proceda a publicar un trabajo y disponibilizar un conjunto de datos referente a ese trabajo, de modo que puedan ser reutilizados, ya tienen una infraestructura de preservación de esos datos, de modo que no sea tan trabajoso a la hora de compartirlos. Existen muchos proyectos, y hay ya muchos conjuntos de datos, que no tienen un gran tamaño, pero sí hay una gran cantidad de ellos; generalmente pertenecen a proyectos menores, a grupos de investigación menores, que no tienen tantos recursos. Para esos investigadores, lo ideal sería preservar los datos de inmediato. Pero también hay proyectos grandes, como los del CERN, que hoy citó Luis, u otro tipo de proyectos con datos médicos, por ejemplo, es importante que estén en repositorios internacionales y grandes, pues de ese modo se pueden preservar y optimizar mejor, además de que pueden ser entregados de acuerdo a la necesidad del usuario. Todo esto es diferente dependiendo del formato de los datos y del área de investigación a que correspondan. Esto muestra también la importancia de los repositorios internacionales dedicados a un área específica, porque uno de los mayores problemas que se enfrentan es la diversidad de datos y formatos. Las infraestructuras deben servir a la mayoría de los investigadores pero muchas veces no se consigue «customizar» lo suficiente para ello, para que la preservación se torne más automatizada. Entonces, concluyendo, los tres niveles son importantes: el internacional, el nacional, porque por ejemplo aquí en Brasil existen algunas cuestiones respecto de colocar los datos en la nube,

porque anteriormente la legislación sólo permitía colocar datos en nubes locales, no internacionales, y asimismo las iniciativas nacionales, sobre todo en términos de recursos, ya que las universidades aquí son federales y muchas veces el gobierno no ve la importancia de destinar recursos a estas iniciativas.

LAUTARO MATAS (Argentina): Con respecto a la pregunta, tomo todo lo que dijeron Luis y Manuela, los tres niveles son importantes, pero voy a patear un poco el tablero con algunas cosas. Sí, esa es la solución ideal, pero no es la solución que vamos a tener, al menos no es la solución que vamos a tener en la mayoría de los países de nuestra región. Si vemos lo que ocurre con los repositorios de literatura, vemos que aún en los repositorios de literatura, mayormente con instalaciones de DSpace, tenemos instalaciones muy antiguas en gran cantidad de instituciones, instituciones que no tienen recursos para mantenerlas actualizadas y que ya es un desafío para muchas áreas, especialmente la de las bibliotecas con sus recursos, mantener eso. Entonces, para ser realista, sería ideal que todas las instituciones puedan tener un repositorio de datos a disposición de sus investigadores para el depósito. Pero eso no va a ocurrir o al menos no va a ocurrir en el mediano plazo. Entonces, los niveles superiores, el nacional y el regional, y quiero hacer un matiz respecto de internacional y regional: digo regional porque por supuesto que depositar hoy en Zenodo es una alternativa a falta de otra cosa, pero es importante que pensemos que al final del día vamos a querer tener muchos esos datos en una infraestructura regional, que todavía no existe y que en un momento voy a comentar un poco qué estamos pensando junto con Red CLARA para intentar construirla. Luego está el nivel nacional y en el nivel nacional, a veces se vuelve a repetir el mismo problema en el nivel institucional. Mantener un repositorio, crear un repositorio de datos para grandes volúmenes de información, como bien dijeron mis compañeros de mesa, es un desafío para muchas disciplinas, y a veces es un desafío que puede exceder las infraestructuras nacionales, incluso de muchos países con los recursos que tienen para investigación, pensemos en los países pequeños y también por los volúmenes de producción que tienen. Y esto con las universidades también pasa: hay instituciones pequeñas que, por el volumen de producción que tienen, hay que ver si tiene

sentido en todos los lugares tener esos dos niveles. Es una discusión que no es mía solamente, sino que la estamos teniendo como países, cuando ellos se ponen a desarrollar en el marco de LA Referencia y nosotros participamos en las discusiones a nivel nacional de algunos países, de qué van a hacer con su infraestructura de datos, es algo que estamos discutiendo: ¿vale la pena empezar con un repositorio nacional o juntar esfuerzos de universidades desde un repositorio que compartan y entonces puedan compartir recursos y tener una infraestructura más confiable? Porque no se olviden que aquí el tema es tener una infraestructura confiable: hay algo peor que no tener esto, que es tener algo malo y que los investigadores no confíen eso o que pierdan sus datos o lo que sea. Eso sería peor que no tenerlo porque nuevamente lo más importante aquí es el cambio cultural en los investigadores para que ellos depositen y reutilicen los datos de su investigación, entonces tenemos que crear infraestructuras confiables, tenemos que ordenar también, a nivel nacional e institucional, el mensaje que le damos a los investigadores y no estar cambiando de estrategias entre medio de uno u otro. La organización al fin tiene que ser nacional, tiene que haber una buena relación entre eso y las decisiones institucionales y tiene que haber políticas nacionales. Nosotros, a nivel regional con LA Referencia y con Red CLARA, estamos intentando organizar toda esta discusión, por un lado para articular las distintas políticas en los distintos países, y aprovechar las experiencias de los que, como Brasil, están avanzando más rápido que otros en este problema, entonces acercar y tratar de hacer redes de colaboración para que haya transferencia de conocimiento de los países que están más avanzados hacia los que están empezando. Y a nivel regional, esta idea que estamos trabajando con Luis y con el equipo de Red CLARA, que es la de ir hacia un ecosistema de ciencia abierta en la región y una parte fundamental de ese ecosistema de ciencia abierta es tener una infraestructura de depósito de datos. Esa estrategia no es para reemplazar los otros niveles, sino que es para hacer una infraestructura regional que esté a disposición de los países y de los investigadores cuando no tengan uno en su institución. La estrategia será trabajar juntos como países para consolidar una infraestructura regional, buscar financiamiento para eso,

que no es algo fácil, porque nosotros no tenemos algo como tiene OpenAIRE o Zenodo, lo mismo con el CERN, que tienen financiamiento europeo y fuentes de financiamiento regional. Nosotros nos tenemos que basar o en financiamientos nacionales en colaboración o como contraparte de financiamiento internacional de otros fondos europeos o del BID, o de otras iniciativas, entonces eso es un trabajo que tenemos que hacer. Paralelamente, por supuesto, discutir y consolidar a nivel nacional las estrategias y ser claro con los investigadores, y en las instituciones trabajar en tratar de ver qué tiene sentido, qué se puede sustentar y qué se puede crear con el tiempo. Siempre tiene que haber una red de contención para el siguiente nivel, es decir, si el nivel institucional no existe, tiene que estar el nivel nacional, y en los países en que ese nivel no existe, tenemos que nosotros trabajar juntos para construirlo a nivel regional y tratar de mantener los datos dentro de nuestra región.

¿Qué alternativas de cooperación se están planteando en su país y para América Latina en su conjunto (tecnológicas, políticas, compras consorciadas de identificadores para datos y metadatos, acuerdos)?

L. E. CADENAS: Sí, hay varios temas que se vinculan con el comentario que hizo previamente Lautaro de las dificultades que tenemos en la región, porque las fuentes de financiamiento no son tan abundantes como lo son en las otras regiones de la Tierra. Sin embargo, Red CLARA ha logrado, por un apoyo sostenido de la Comisión Europea desde su fundación, construir una infraestructura muy relevante en términos de magnitud, con conectividad entre los países, que constituye una base importante para poder generar esta integración regional. LA Referencia, por su parte, tiene la representación institucional de los gobiernos que participan en la red y esos son espacios que efectivamente creo que podemos utilizar para promover iniciativas que sean colectivas y regionales. Creo que es lo más difícil de hacer: a nivel nacional, por supuesto, dependiendo de los gobiernos y de las orientaciones y de la percepción que se tiene de la ciencia y de la ciencia abierta puede que cambien las políticas y sean muy diversas. Creo que a nivel regional, las organizaciones que están trabajando en esto están bastante claras y tenemos la oportunidad,

porque tenemos infraestructura construida que permite efectivamente tener repositorios de datos que pueden ser usados por muchos países a nivel regional y que puedan conectar estos datos con datos en otras partes del mundo. Recientemente, ustedes podrán haber visto que inauguramos la conexión directa que ahora tenemos con Europa con el cable submarino, y eso no va a dar una posibilidad de integración muy importante también con las iniciativas europeas en temas de ciencia abierta. Creo que es una labor que va tomando tiempo, pero quizás podamos conseguir hacer confluir los intereses y las voluntades no solo de los gobiernos, sino de las agencias internacionales de cooperación, que están siendo muy importantes en términos de impulsar esta visión, esta necesidad de contar con datos abiertos y de contar con una visión de ciencia abierta a nivel global. Eso ustedes lo conocen mucho mejor que yo, pero todo lo que tiene que ver con la resolución de la UNESCO alrededor del tema de ciencia abierta, que esperamos que esté refrendada (que no sé ni siquiera si ya está refrendada o no, creo que no, pero debe estar por estarlo), eso genera un marco institucional y un marco político muy relevante a nivel global, con una cantidad enorme de países respaldando la necesidad de tener este enfoque de ciencia abierta. En los cuatro pilares planteados por el documento de la UNESCO, está el tema de las infraestructuras, que es un tema que tenemos que reforzar enormemente. Lautaro mencionaba la necesidad de que los datos estén en la región y yo coincido plenamente. Recientemente, la UDAT publicó su reporte sobre economía de este año, un reporte muy interesante que refleja, principalmente, todo lo que tiene que ver con los datos y de qué manera los datos están siendo dominados o manejados por básicamente cuatro o cinco empresas a nivel global, que tienen toda la capacidad de almacenamiento, de gestión, de procesamiento y yo creo que hay que lograr un equilibrio, ningún desequilibrio es sano; yo creo que hay espacio para los servicios de nube privados o comerciales, pero también es necesario tener políticas de datos en los países, que garanticen que los datos propios son manejados de una forma adecuada. En ese mismo reporte se hablaba de que el 95 % de la innovación de las *startups* en inteligencia artificial están en Estados Unidos y en China. El otro 5 % está en el resto del mundo incluyendo muchos

otros países que son desarrollados, pero que no tienen el mismo nivel de desarrollo en términos de esta tecnología. Eso va a generar un gran desbalance y vamos a necesitar un enfoque global y un enfoque regional alrededor del tema. Red CLARA participa en muchas instancias de cooperación, obviamente las instancias que tienen que ver con cooperación con Europa, participamos en instancias que tienen que ver con la cooperación en redes de investigación y educación a nivel global, estamos hablando de más de 140 redes; participamos en organizaciones donde podemos llevar este mensaje, esta necesidad y está idea de poder construir colectivamente un espacio de datos para los investigadores, no sólo a nivel de Latinoamérica, sino a nivel global. Yo creo que hay que tomar todas esas instancias, que están allí, que están además trabajando colectivamente para lograr manipular cada vez más datos enormes. Por ejemplo, en el ámbito del cambio climático: es un ámbito que está permitiendo organizar mucho, cómo estructurar y manejar datos en una determinada área, por la relevancia que esto tiene para temas de monitoreo la Tierra o para tomar decisiones. Todos los esfuerzos que hay al nivel de Naciones Unidas en términos de los datos, de las estadísticas y de las informaciones, etc. Creo que hay que juntar todo esos esfuerzos y buscar los espacios necesarios. Ahí está la disposición de las organizaciones que tenemos en la región: LA Referencia es un mecanismo enorme y muy importante para coordinar las acciones políticas o por lo menos plantear ideas en el marco de acciones políticas regionales dentro de los espacios, los ministerios de ciencia y tecnología en la región y tenemos que aprovecharlos todos, porque realmente el reto es extremadamente grande, los recursos son pocos, y por eso es indispensable que tengamos una estrategia colaborativa, una estrategia donde nos sumemos unos con otros y podamos complementar lo que no tenemos a través de financiamiento estructurado a nivel regional. Hay que buscar todos los espacios abiertos para esta conversación, los distintos grupos que existen; por ejemplo, participé hace poco en un evento de la Alianza del Pacífico: ese es un ámbito en el cual hay que plantear el tema y decir hace falta infraestructura, hace falta financiamiento, hace falta una visión común en el marco de todas las instancias, tanto políticas como técnicas como de todos

los niveles que tenemos en la región, hay que llevar este mensaje y defender la idea de la ciencia abierta y la importancia que tiene para nuestra región y la relevancia que tiene que tengamos dominio sobre nuestros datos, que decidamos qué podemos poner donde queramos y qué queremos tener nosotros, porque es indispensable para nuestro desarrollo.

M. KLANOVICZ FERREIRA: Luis mencionó varios puntos relacionados con la Red CLARA, y creo que estas iniciativas son muy importantes porque hay instituciones que tal vez no tienen capacidades o infraestructura y hay que conseguir promover la preservación de los datos localmente, para tener cierta uniformización entre países distintos, así como promover la importancia de la ciencia abierta y promover incentivos en esa área. Aquí en Brasil, existen algunas iniciativas de cooperación, en las cuales estamos participando y procurando recursos pero el objetivo de uno de los consorcios es la infraestructura para lograr un ecosistema de investigación colectiva. Inicialmente, este consorcio sería utilizado para comprar DOI, a partir de Data Cite, porque de este modo existen algunos descuentos importantes. Este sería un primer paso para incentivar la preservación de los datos. Esto sería sólo el comienzo, la idea es que haya más asociaciones colaborativas que compartan sus recursos. Otra iniciativa que existe en Brasil ha hecho un relevamiento de la situación de los los repositorios de datos del país y de sus necesidades. A partir de eso este año fue lanzado un plan para las instituciones que no tienen infraestructura suficiente y durante un período de incubación de nueve meses se les brinda capacitación y transferencia de recursos. Estas son dos iniciativas nacionales que veo muy importantes y necesarias para que se dé la voz de largada de los repositorios de datos; todo esto es fundamental para la comunicación de la ciencia, aunque a veces se vea como una «obligación» más para los investigadores, ya que deben también depositar los datos. Es decir, que estas iniciativas son importantes para que haya una concientización de la relevancia de los datos, de cómo teniendo los datos disponibles de una investigación anterior, teniendo datos confiables y pudiendo reproducirla, así se evolucionará más rápido y podremos llegar, como sociedad, en Brasil y en el mundo, a nuevos descubrimientos útiles. Esa concientización entonces es muy

importante y estas iniciativas, en mi opinión, sirven no sólo para brindar infraestructura sino para esto, si todo el mundo lo está haciendo, por qué no vamos a hacerlo; es preciso convencer a los dirigentes de que esta es una aplicación adecuada de los escasos recursos con que se cuenta.

L. MATAS: Bueno, lo bueno y lo malo de hablar después de gente tan completa es que quedan pocas cosas que decir. Hablando en serio, más allá de los puntos específicos en los que estoy de acuerdo, aquí es una actuación conjunta la que estamos haciendo. Entonces muchas de las cosas que yo diría, son las que ha mencionado Luis. Lo que quiero remarcar es que, como decía antes, sí, estamos muy activos buscando esta integración regional y estamos complementando la experiencia de Red CLARA en los proyectos y también el capital simbólico que tiene Red CLARA frente a las iniciativas europeas y además con las grandes inversiones que han hecho en términos de conectividad en la región, con este rol que tenemos nosotros es de poder poner en una mesa los organismos de ciencia y tecnología en temas de ciencia abierta a discutir y a consensuar políticas y también estrategias y ojalá a poder consensuar también con otras iniciativas de la región en trabajar juntos en un proyecto que nos permita crear y financiar estas infraestructuras. Nosotros tenemos que crear artificialmente esto que otras regiones tienen como dado, que es el financiamiento conjunto para ser usado en este tipo de iniciativas. Esta articulación que estamos haciendo con Red CLARA desde lo cooperativo, que para nosotros es muy fácil porque estamos muy relacionados y confiamos mucho en esa experiencia, pero, por otra parte, la opinión de otros países y esa mesa de discusión esperamos que nos permita crear uno o varios proyectos lo suficientemente robustos para poder conseguir el financiamiento que necesitamos. Sobre todo el segundo desafío que viene, porque a veces hablamos de financiamiento para conseguir montarlo, pero el problema en sí no es montarlo, el problema es sostenerlo. Ahí es donde está el mayor desafío a largo plazo porque nosotros, con suerte, podríamos montar una infraestructura compartida o distribuida en varios nodos, de distintas cosas, que para mí sería lo ideal, tratar de hacer replicación y crear algo que sea realmente distribuido, para que la preservación sea confiable y, de nuevo, este tema de darles a los

investigadores algo confiable donde depositar. Estuvimos discutiendo la posibilidad de hacer un Zenodo regional, que se pueda hacer algo en forma distribuida, y sea algo parecido a lo que es Zenodo; podríamos solucionar el problema de la soberanía de los datos que en algunos países, como decía Manuela, en términos de las legislaciones obligan a tener los datos, eso se puede así, pero replicado en otros lugares para seguridad. El problema es como sostenemos: hay que comprometer a los Estados en darle sustentabilidad a eso y yo creo que ahí el rol de LA Referencia, el rol de interlocución que podemos tener, tiene que ir en ese sentido, tiene que ir en el sentido de convencer, de convocar y de generar un ámbito, un fondo para luego mantener esto funcionando. Es un trabajo a muy largo plazo, un trabajo que ya estamos haciendo, pero que nos va a llevar tiempo y que es un trabajo siempre de mediano y largo plazo. En el ámbito de los datos, el tema de la infraestructura nos va a ocupar los próximos cinco o diez años de consolidar y llegar a un punto comparable al que otras regiones están en estos términos. Pero hay que hacerlo y lo estamos haciendo ya.

¿Cómo se deben pensar las plataformas de publicación para integrarlas con infraestructuras de datos? O sea, ¿cómo se adaptan los circuitos de publicación, qué protocolos de interoperabilidad se utilizan, qué servicios deben brindar los sistemas de publicación, etc.?

L. E. CADENAS: Yo tengo que hacer un *disclosure* aquí: de eso conozco muy poco realmente, con lo cual no me siento competente para responder apropiadamente. Sí puedo decir, en términos de los servicios, que mucho de lo que se está desarrollando en el marco de estas iniciativas que mencioné previamente, es lo que hace falta: tener garantías de que el investigador pueda manejar el ciclo completo de los datos y vincularlos con sus publicaciones. Creo que hay elementos importantes que considerar, por ejemplo la seguridad alrededor de los datos por una parte, y por la otra, la integridad y mantenimiento de esos datos, de manera que haya una garantía de que lo se publica y los datos en los cuales se basa la publicación sean preservados adecuadamente en el tiempo y no cambien y puedan ser reproducibles, a los efectos de esa

capacidad de reproducibilidad que queremos buscar a partir de este tipo de herramientas. También apoyar el manejo de las aplicaciones o los software para el desarrollo y para el procesamiento de esos datos es un elemento que es importante y la posibilidad para el investigador de definir los niveles de apertura y la forma en que desea manejar sus datos en el marco de las publicaciones que está proveyendo.

M. KLANOVICZ FERREIRA: Uno de los grandes problemas en relación con los datos, su reuso y la reproductibilidad de las investigaciones es que los datos estén documentados de forma adecuada. En la actualidad, existen los **principios FAIR**, que tienen como objetivo volver los datos encontrables, accesibles, interoperables y reutilizables. A partir de estos objetivos, los repositorios de datos deben implementar acciones para que sean alcanzados. Una de las formas para hacer eso, y que forma parte de mi investigación, por ejemplo, es la documentación automatizada de datos. Documentación relacionada a los datos o bien relacionada a las investigaciones en las que se obtuvieron esos datos. En la actualidad, gran parte de las investigaciones son llevadas a cabo mediante computadoras, pero no todo el mundo tiene la capacidad de describir las herramientas con las que se obtuvieron los datos o el modo en que un experimento fue ejecutado, y asimismo documentar la procedencia de los datos, qué días fueron tomados, en qué horario, qué entidades estuvieron involucradas en su recolección, ¿son confiables estos datos? Hoy día existen iniciativas en proyectos grandes, como en el CERN, que disponen de herramientas para que toda esta información sea guardada junto con los datos. Estas son herramientas que le facilitan la tarea al investigador, porque si un dato no fue bien documentado, si su procedencia no está documentada, otros investigadores no lo van a poder utilizar. Entonces es importante proveer de estas herramientas a los investigadores que hoy tienen dificultades para documentar correctamente los datos. Otra cosa importante es incentivar, y desde luego que todo depende de los recursos, la preservación de los datos dentro de la publicación. Muchas veces los investigadores necesitan publicar los datos porque se los exigen para poder enviar el trabajo a una publicación. ¿Por qué los editores quieren los datos? Porque así pueden

revisarlos antes de la publicación y esa revisión, siempre que sea posible, aunque es difícil, es muy importante, y teniéndolos bien documentados se facilita mucho más esa revisión. Esa revisión incentiva, a su vez, la calidad de los datos, siempre teniendo presentes los principios FAIR. No es sólo una cuestión de verificación de la investigación, sino que es una cuestión de incentivos a la investigación. Un tercer punto que me gustaría comentar, muy ligado a la pregunta anterior, es la evaluación de los datos, así los investigadores podrían ver cómo los evalúan los usuarios; además de eso, los accesos a los datos, accesos a los documentos, descargas, de forma que el usuario que vaya a usar esos datos consiga ver rápidamente si el dato ya fue usado, si tiene muchos *downloads*, cuántos accesos tiene, si fue calificado, etc. Conforme la utilización de los datos aumenta, se debe facilitar a los usuarios toda esta información contextual y así habrá también un aumento en la compartición de los datos, que puede ser exponencial. Este tipo de iniciativas son, entonces, muy interesantes y en mi opinión deben ser incentivadas por los repositorios.

L. MATAS: Creo que la respuesta de Manuela fue muy buena y interesante y abarca algunas cosas que no pensaba mencionar y que puedo entonces mencionar. En principio, lo importante es la calidad de los metadatos y en este sentido los nuevos estándares, ya no son tan nuevos, pero nosotros todavía estamos empujando para que se implementen, que son los estándares de metadatos OpenAIRE 4, que tienen contemplado poder vincular los datos con las publicaciones, por medio de los identificadores relacionados, especificando que eso es un dataset. Por eso es ideal que donde depositamos los datos, tengan identificadores únicos y así poder citarlos. La calidad, entonces, para que todos esos metadatos sean encontrables, reutilizables, la calidad con la que son depositados, los lugares donde son depositados y cuán bien curatos están esos datos y también el orden que tienen los investigadores para poder después en las publicaciones citarlos correctamente y vincularlos. En ese sentido, lo que nosotros tenemos en LA Referencia, es un grupo de trabajo, como decía antes, para intentar que todas esas implementaciones, la migración y la migración hacia atrás, que no es nada fácil, ocurra, para todo eso lo que

nosotros tenemos es la nueva versión del cosechador; también estamos trabajando en la posibilidad de hacer migraciones entre esquemas de metadatos de lo que tenemos de herencia en OpenAIRE 3 para OpenAIRE 4. Queremos poder propiciar eso y poder dar un servicio a los repositorios, darles una versión enriquecida de los datos en otro esquema que no tienen para que no hagan ese esfuerzo manualmente. Eso es un sueño que tenemos y la tecnología que tenemos y que hemos desarrollado en los últimos dos años nos lo permite hacer, hay que organizarlo y lo vamos a hacer. Por otro lado, me disparó una de las cuestiones que mencionó Manuela, la relación con un proyecto en el que estamos trabajando y mirando con mucha atención con Luis en el ámbito de Red CLARA también y con otros socios que es el proyecto de MyLab: que los investigadores y los grupos de investigación tengan un ambiente donde trabajar ordenadamente y que queden registrados los contextos donde la investigación es hecha. En las presentaciones de LA Referencia yo a veces digo que nosotros empezamos ocupándonos de la literatura, de los repositorios, pero que consideramos como producción científica no solamente a la literatura y los datos, sino a toda la información contextual que forma parte de los procesos de investigación y de generación de ese conocimiento. Entonces me pareció muy bueno lo que dijo Manuela, porque la reproductibilidad, que es la condición necesaria para ser ciencia, no es solamente tener el artículo y tener los datos, sino tener el contexto en que ese experimento o esos datos fueron generados. Lo de MyLab nos lo acerca, en el sentido de que ellos están trabajando en una herramienta para que los grupos de investigación puedan preservar hasta las discusiones que ocurren en el proceso de investigación. Lo tomo hoy, porque me parece que es algo interesante para profundizar y para poner muy de relevancia, en nuestro trabajo del próximo tiempo. Eso es lo que estamos haciendo, son muchas cosas y son muchos frentes, pero estamos trabajando un poquito en cada uno y tratando de avanzar como comunidad, en esto lo importante es avanzar sin prisa y sin pausa.

Mesa 5: Experiencias de nodos nacionales de LA Referencia

WASHINGTON RIBEIRO SEGUNDO

Instituto Brasileiro de Informação em Ciência e Tecnologia – IBICT (Brasil)

JUAN MALDINI

Gerente de Servicios Digitales en Agencia Nacional de Investigación e Innovación (ANII) (Uruguay)

PAOLA A. AZRILEVICH

Coordinadora Biblioteca Electrónica de Ciencia y Tecnología, MinCyT (Argentina)

ANDREA M. MORA CAMPOS

Sub Comisión de Conocimiento Abierto de CONARE (Costa Rica)

PAOLA CASTRO MONROY

Red Colombiana de Información Científica – REDCOL (Colombia)

FREDDY SUMBA

Especialista en desarrollo de servicios académicos de CEDIA (Ecuador)

Modera: **LAUTARO MATAS**

Secretario Ejecutivo y Técnico de LA Referencia (Argentina)

RESUMEN EXTENDIDO

LAUTARO MATAS (Argentina): Voy a comenzar dando un pequeño contexto de LA Referencia para que nos sirva como introducción a aquellos que se conectan y que no conocen el trabajo de LA Referencia. Nosotros somos la red de repositorios de ciencia abierta, creada en el año 2012 por un acuerdo de gobiernos; esta red fue creada luego de un proyecto BID que estuvo gestionado

por RedCLARA e integrado por varios gobiernos de la región, con la idea de justamente poner en marcha un plan de de repositorios institucionales que luego tomó el carácter de federada. Desde el principio y en la discusión ya del proyecto se planteó que íbamos a hacer esta red, a diferencia de otras regiones, con una idea de nodos nacionales fuertes, donde sean éstos quienes cosechan y decidan, administren y gestionen cómo los contenidos de los repositorios iban a ser integrados y luego esa colección se integraría en una colección regional. Entonces nuestra misión principal es dar visibilidad a la producción, ya sea los repositorios, pues esto que empezó con repositorios de literatura y con el acceso abierto en los repositorios en aquel momento, ha devenido en algo mucho más amplio, que es la ciencia abierta: ahora hay repositorios de datos y entendemos ya a la producción científica como algo bastante más amplio. Ayer hablábamos la mesa 4 que no son solamente los productos de investigación o los resultados de investigación sino que también tenemos que pensar en poder dar visibilidad a los procesos y a los contextos en los que esos productos son generados. De todo eso nos ocupamos; fundamentalmente, mediante una serie de acuerdos políticos, LA Referencia es, en realidad, una gran mesa de discusión permanente y de decisión integrada por los representantes de los países que nos forman. Hoy tenemos precisamente a varios de los representantes en esta mesa de debate. Estas personas o varias de estas personas son las que integran el consejo directivo de LA Referencia, que es nuestro órgano de gobernanza, donde se toman todas las decisiones en forma democrática, y se fijan también los rumbos y las visiones que nosotros tenemos como región frente también a otras regiones del planeta. Para entender un poco las preguntas que van a venir a continuación, vuelvo a hacer énfasis en que lo que nosotros hacemos desde el punto de vista técnico, además del tema político, es tener una plataforma de software que funciona en los nodos nacionales de LA Referencia y en el nodo regional, y que permite cosechar, validar, de acuerdo a directrices, estándares e indizaciones internacionales, enriquecer los metadatos que son cosechados de los repositorios, poder agregarlos a nivel nacional, generar portales de búsqueda para visibilizar esa producción a nivel regional y a nivel internacional también.

Todos estos metadatos son cosechados por otras iniciativas, como por ejemplo OpenAIRE. Hoy tenemos la suerte de contar con varios de los consejeros e integrantes de los equipos técnicos también, de los nodos nacionales de LA Referencia, que son quienes van a intentar entre todos contestar algunas preguntas sobre cómo será el proceso de esta integración y de agregación de repositorios.

¿Cuáles han sido los principales retos o dificultades que han afrontado para lograr la interoperabilidad de los repositorios de su país con LA Referencia y qué estrategias han implementado para resolverlas? ¿Cuáles consideran que han sido los mayores beneficios y áreas de oportunidad para ustedes como responsables técnicos y para las instituciones de sus países como miembros de LA Referencia?

PAOLA AZRILEVICH (Argentina): Para nosotros siempre este tema de interoperabilidad estuvo asociado con la calidad de los repositorios. Entendemos que no hay repositorios de calidad si no cumplen ciertas condiciones y reglas de interoperabilidad. Los repositorios tienen que ser interoperables, tienen que permitir su integración con otras redes de repositorios y el intercambio de información, y para ello desde que creamos el Sistema Nacional de Repositorios (SNRD) en 2011 trabajamos en definir directrices y pautas que reglamenten estas cuestiones. En el 2013 cuando comenzó efectivamente a funcionar el SNRD definimos las directrices de ese año que tenían quince elementos DublinCore no cualificado e incluía en la exigencia del uso de los vocabularios Driver para tipo de recursos e incluso hicimos traducciones de esos tipos de recursos en un campo aparte para no afectar la interoperabilidad a las traducciones al español; también usábamos los tipos de versiones de Driver y listas de términos aceptados (por ejemplo, ISO para los idiomas, etc.). Estas directrices, si vamos a hablar de interoperabilidad, no pueden ser sólo hacia adentro, estaban basadas en distintas directrices como las Driver 2.0 y las OpenAIRE 1 en ese momento, más algunos modelos que habíamos analizado con BDCOL (Colombia) y también una propuesta que había en la Universidad de Cuyo, a nivel local. En ese momento no había mucho más que esto y ya ahí planteábamos que

evaluaríamos las nuevas directrices que se establecieran en el marco de lo que entonces era el proyecto BID que culminó siendo LA Referencia, y también las nuevas directrices de OpenAIRE que se estaban desarrollando. Eso sucedió por lo tanto hacia el 2015, momento en que hicimos una nueva versión de las directrices que establecieron con veinticuatro elementos DublinCore no cualificado, seguíamos con los vocabularios Driver con traducción al SNRD y también, como indicaba antes, para tipo de versiones, para tipo de recursos y listas de términos aceptados. En esta oportunidad ya nos basamos directamente en la OpenAIRE 3.0 y en los metadatos y políticas de cosecha de LA Referencia, que ya había lanzado su documento de criterios por los cuales aceptaba o dejaba de aceptar determinados registros que estaban en los repositorios y por ende en los nodos nacionales. En la actualidad, estamos terminando de elaborar las directrices que estimo serán lanzadas a comienzos del 2022 con 32 elementos implementando los vocabularios COAR, que son multilingües ya dejamos de hacer traducciones propias y locales para tipos recursos, tipo de versiones, tipos de acceso más listas de términos aceptados en distintos elementos que siguen siendo requeridas. Estas directrices estarán basadas en las directrices de OpenAIRE 4.0 y también en el último esquema de DataCite. Esto nos permitirá describir tanto publicaciones como datos de investigación con una sola directriz y dependerá la obligatoriedad de cada elemento, de acuerdo al tipo de recurso del que se trate. Creo que en todos estos años ha sido crucial nuestra participación en los distintos grupos que han abordado estas cuestiones a nivel regional e internacional: trabajamos en el desarrollo de los metadatos y políticas de cosecha de LA Referencia, trabajamos en los grupos de vocabularios controlados de COAR y también trabajamos con OpenAIRE en la elaboración de sus nuevas directrices, con lo cual todo lo que bajamos a nivel nacional contempla claramente lo internacional porque hemos participado en su colaboración. Me parece que esto es crucial así como es crucial la autonomía que le hemos dado a los repositorios argentinos en conocer sus diagnósticos, saber en qué fallan y en qué tienen que mejorar en cuanto al cumplimiento de estas pautas. Hasta el 2016-2017 hemos trabajado con DNET y un desarrollo propio que le

encargamos al SIU para diagnóstico que le llamamos «Observador del SNRD», donde los repositorios podían ingresar y ver directamente sus diagnósticos, ver en qué estaban fallando y que debían mejorar para que los registros fueran aceptados. Nosotros cuando hay registros obligatorios y elementos obligatorios que no se cumplen, el registro del repositorio en el SNRD queda rechazado y creo que eso es importante decirlo también. Hacia el 2016-2017 e incorporamos el software de LA Referencia, con lo cual tenemos la posibilidad de editar los validadores, de editar transformadores y esto nos permite tener distintos niveles de transformación de acuerdo a la evolución de los repositorios. La tendencia es transformar lo menos posible y exigir que los repositorios cumplan de base con los requerimientos, pero a veces tenemos problemas frecuentes, como que en lugar de usar el ISO de tres dígitos usar el ISO de dos y eso no nos parece tan grave como para no transformarlo pero sí les avisamos para que puedan corregirlo. Con el software de LA Referencia tenemos la posibilidad de acceder a los diagnósticos; los que no pueden acceder todavía son los repositorios, por ende hemos realizado algunas estrategias que son exportar los resultados de las validaciones y enviarlos a los repositorios. Trabajamos mucho en esto, en un seguimiento día a día de cada cosecha, que tiene distintas instancias de validación, con transformaciones y sin transformaciones y le hacemos llegar los dos diagnósticos para que vean la diferencia: qué pasa si nosotros no transformáramos y qué pasa cuando los cosechan otras redes que no pasan por nosotros, por ejemplo. También lo que logramos es una serie de visualizaciones en Tableau Public que les permite también hacer ese seguimiento histórico y ver cuáles son sus fallos. Hoy en día en Tableau tenemos unas 148 visualizaciones de cada uno de los repositorios donde pueden filtrar por los errores, cuáles son los obligatorios, cuáles son los optativos, en qué están fallando en este sentido y cuáles son los registros que están bien y válidos para poder corregir. En cuanto a la segunda pregunta, por un lado tuvimos la posibilidad de participar de todas estas instancias que decía antes y mejorar lo que estamos haciendo, trabajar en las directrices, trabajar en los grupos de vocabularios internacionales, de poder hacer una bajada muy de primera mano hacia los repositorios de todo esto. Las herramientas de

diagnóstico que nos da LA Referencia de validación y transformación son realmente útiles e interesantes y tienen la flexibilidad de poder tener distintas instancias de validación. Creo que participar de estas instancias a nivel regional, ya no sólo desde lo técnico sino desde lo político, porque si no si lo técnico sin lo político no nos permitiría avanzar del modo que avanzamos, y me parece que tener el apoyo de las autoridades y de LA Referencia como corpus nos hace mejorar como región y nos hace poder empujar más rápido el carro. También las posibilidades de generar instancias de capacitación: a nivel regional hemos tenido alguna experiencia en el marco de LA referencia con capacitaciones a los repositorios. Este año hubo una serie de webinars que son súper interesantes para los nodos y también para los repositorios. A la vez, creo que todavía LA Referencia no desarrolló su máximo potencial en eso, llegar a los repositorios con capacitaciones a instancias de mejorar en cuanto a software metadatos, etc., si bien se ha canalizado a través de los nodos me parece que el potencial que tenemos es sumamente fuerte y muy importante.

PAOLA CASTRO MONROY (Colombia): Desde el nodo Colombia, queremos primero plantear algunos antecedentes de todo lo que es esta iniciativa, algunos inconvenientes y problemáticas que vivimos para lograr todo lo que es la visibilización y la recuperación de la información científica a nivel nacional para dar luego lo que son los logros y los beneficios que nos trae todo este proceso con LA Referencia. Entre los antecedentes, hacia el año 2007 fue cuando se crea la Biblioteca Digital de Colombia (BDCOL), que surge a partir de un proyecto de tres universidades nacionales y el Ministerio de Educación nacional, y posteriormente nace lo que es el primer Portal BDCOL: la idea fue implementar una infraestructura tecnológica que permitiera tener acceso libre y gratuito a los productos de investigación. Es allí, cuando contamos con esa infraestructura tecnológica, que LA Referencia nos hace una invitación a ser parte de toda esta estrategia y es cuando nuevamente entre el Ministerio de Educación nacional, el Ministerio de Ciencias, RENATA, y las instituciones vinculadas a BDCOL, se crea el Sistema Nacional de Acceso Abierto al Conocimiento (SNAAC). Se fortalece esa estrategia en su momento de acceso al conocimiento científico, pero ya en 2017 se plantea el redimensionamiento

del CENDOC, que es lo que hace que a partir de todo este estudio nazca la Red Colombiana de Información Científica. Podemos ver algunas de la problemáticas que fuimos encontrando cuando teníamos el SNAAC, por ejemplo, que la plataforma no contaba con una actualización; durante dos años la vinculación y la cosecha de las instituciones se había detenido; no todos los productos que se estaban presentando correspondían a los productos que estaban dentro del modelo de medición que tiene el Ministerio de Ciencia y no había una visibilidad tampoco de los productos de investigación financiada con recursos públicos; no había estándares, tampoco teníamos un estándar claro para la construcción de repositorios institucionales y no había lineamientos claros frente a la propiedad intelectual, las infraestructuras tecnológicas y la gestión y recuperación de la información científica que tenían las instituciones. Esto conllevó a que Colombia no estuviera realizando nuevas cosechas y vinculación de instituciones hacia LA Referencia. Entonces a partir del CEDOC nace la Red Colombiana de Información Científica en el 2019, con el propósito de articular a todos los expertos y actores del sistema de tecnología e innovación para potenciar y dar visibilidad a la producción científica colombiana. Se establecen dos estrategias principales o bien dos fases: la primera fue desarrollar la infraestructura tecnológica que nos permitiera visibilizar toda esta producción científica y es aquí cuando nos apoyamos en la infraestructura tecnológica que nos da LA Referencia. Luego continuamos con un proceso de vinculación de instituciones a nivel nacional entre universidades, centros, institutos de investigación con un proceso más formal y con un mapeo que se hizo frente a todas las instituciones que tenían la mayor capacidad de producción científica. Para pertenecer a la Red desde el nodo Colombia se planteó que deben presentar ciertos requisitos: primero contar con un repositorio institucional y contar con producción científica que estuviera alineada con los productos de investigación del modelo de medición de grupos y contar con los lineamientos para poder dar calidad a la producción científica que se recuperaba allí. Adicionalmente se planteó un fortalecimiento de capacidades a nivel nacional a todos los equipos de las instituciones, principalmente en gestión y recuperación de la información científica y es allí

cuando se realiza la primera cosecha de las instituciones vinculadas, que en esta primera fase, como mencionaba, se hizo un mapeo y se logró 30 instituciones en un principio. La fase 2 fue fortalecer, de acuerdo las problemáticas que les mencionaba. y es cuando la Red Colombiana de Información Científica crea las directrices para repositorios de investigación de la Red: estas directrices se hicieron con base en las OpenAIRE 4 y la idea era hacer una extracción muy profunda y presentárselas a todas las instituciones y en todo lo que queríamos hacer a nivel nacional con los repositorios institucionales para la calidad de la información, adicionalmente también se encuentra todo el tema de los vocabularios COAR y Data Cite, pero lo que se quiso también fue complementarlas con vocabularios: allí es cuando también se incluyen los vocabularios de RedCol para que se complementara y fuera acorde a las 58 tipologías del modelo de medición de grupos. Una vez creadas las directrices, empezamos un proceso de fortalecimiento de capacidades y de capacitación nuevamente con todos los equipos de las instituciones, se hace un proceso de fortalecimiento y las instituciones vuelven y ya tienen más claro todo este tema. Posteriormente a cada institución vinculada se le genera un diagnóstico; es decir, se hacen tres diagnósticos principales: el primero es evaluar el repositorio como sistema de información, se evalúa todo el tema de política, aspecto legales, interoperabilidad, estadísticas, seguridad; luego se evalúa la calidad de metadatos, como todo en el repositorio institucional aquí se evaluó la taxonomía, los vocabularios utilizados, las tipologías, las licencias, etc. Se le dio un enlace en un *dashboard* a cada institución para que entrara y pudieran mirar el detalle de todo lo que tenían en sus repositorios y, por último, lo que se hizo fue evaluar el cumplimiento de los criterios de acuerdo con LA Referencia, con base en las cuatro tipologías que en su momento se estaban manejando; se hizo una revisión de cuáles eran los campos obligatorios, cuáles eran los campos opcionales, las recomendaciones y todo eso para lograr la recuperación y validación de sus productos. Luego, se hace un proceso de seguimiento y acompañamiento de las instituciones: se ha generado una hoja de ruta con ellos y se hace continuamente acompañamiento. En el marco de la pandemia también se hicieron, aprovechando el tema de la virtualidad, y se ha

trabajado y se ha avanzado claramente en todo este proceso a la calidad de la información. Los resultados: a la fecha tenemos 68 instituciones vinculadas a la Red, entre instituciones y centros de investigación, también tenemos 50 repositorios que están leídos desde el recolector nacional y posterior al recolector de LA Referencia contamos con más de 144.000 productos de investigación en acceso abierto. Hemos hecho el acompañamiento y fortalecimiento de los repositorios de las instituciones miembro y también en gestión y recuperación del equipo humano con el que cuenta cada institución. Hemos publicado las directrices para repositorios institucionales y ya venimos trabajando también en la nueva de repositorios para datos de investigación. Todo este proceso que ha hecho RedCol le han permitido ser hoy día a la Red Colombiana de Información Científica el nodo nacional entre la Red y LA Referencia. Por último, mencionar rápidamente los beneficios de la articulación y fortalecimiento de servicios y lineamientos para lograr la visibilidad la información científica; también la conexión internacional de las estrategias que nos ha permitido por medio de LA Referencia con COAR dar la visibilidad de la producción científica a nivel internacional y a desarrollar capacidades y proyectos alineados con los otros modelos internacionales.

JUAN MALDINI (Uruguay): Uruguay es uno de los nodos más recientemente incorporados a LA Referencia. Nuestro sistema nacional de repositorios se llama SILO y se creó en 2019 coincidentemente con el ingreso a LA Referencia. Hoy cosecha siete repositorios con un total de 16.500 registros, de los cuales unos 6 mil y pico son cosechados por LA Referencia y estamos trabajando con tres o cuatro instituciones más que están desarrollando sus repositorios o ajustándolos para poder interoperar. El sistema de ciencia y tecnología uruguayo es chico y estaba muy centralizado en algunas instituciones, principalmente la Universidad de la República (UDELAR), que es la principal universidad pública. Antes de responder las preguntas específicas que nos proponen, a mí me gustaría brevemente repasar cuál era el estado de los repositorios en Uruguay antes de 2019; es decir, antes de la adhesión a LA Referencia y de la creación de SILO. En Uruguay no tenemos una política de acceso abierto de alcance nacional, no hay legislación al respecto. En ese

momento había tres o cuatro repositorios institucionales, ninguno de ellos interoperable e incluso no tenían levantado sus servidores OAI; solamente en un caso tenían una política de la institución explícita de acceso abierto y en general no existían mandatos de autoarchivo en los repositorios o cuando existían no se cumplían por falta de incentivos o de control, problema que todavía persiste en muchos casos. Incluso para identificar la producción nacional o institucional era necesario recurrir a las bases de datos bibliográficas internacionales con todos los sesgos conocidos. Entonces empezamos un trabajo primero de difusión y concientización a nivel de las autoridades institucionales e investigadores sobre la importancia de implementar y consolidar los repositorios institucionales, de desarrollar políticas de acceso abierto y en todo este proceso el apoyo de LA Referencia y el intercambio con los países miembro de LA Referencia fue fundamental. Dado este panorama, paso a tratar de responder las preguntas que se proponen. En relación a los desafíos sobre interoperabilidad, cuando este proceso empezó, no teníamos ningún ningún repositorio cumpliendo estándares de interoperabilidad, así que se hizo un trabajo de estandarización en base a las directrices OpenAIRE 3 y el documento de metadatos y políticas de cosecha de LA Referencia; hubo un trabajo por parte de las instituciones de normalización de los tipos documentales y de inclusión o normalización de las versiones según los vocabularios propuestos por OpenAIRE 3, pero como sabíamos que se venían las directrices 4.0 de OpenAIRE, lo que hicimos fue posponer la definición de las directrices de metadatos propias del nodo y por lo tanto su adopción a nivel de los repositorios. Varios repositorios decidieron ajustarse a las directrices OpenAIRE 3, pero en general la estrategia fue implementar transformaciones a nivel del cosechador de SILO, para estandarizar los metadatos mientras esperábamos el lanzamiento de las OpenAIRE 4 y la elaboración de nuestras propias directrices en función de esto, que es un proceso que recién estamos empezando. Entonces la implementación del cosechador de LA Referencia y el traspaso del conocimiento en relación a él son fundamentales para lograr rápidamente la interoperabilidad, no necesariamente de los repositorios directamente pero por lo menos sí desde el

nodo nacional. En cuanto a la segunda pregunta, los mayores beneficios y áreas de oportunidad para las instituciones, es difícil disociar como ya comenté el proceso de creación de SILO como sistema nacional de repositorios a la adhesión a LA Referencia, porque LA Referencia ofreció infraestructura tecnológica, un marco institucional y, sobre todo, modelos nacionales para tomar como referencia a la hora de crear el sistema. Así, fue fundamental en todo el proceso. A nivel de las instituciones, lo que ellas mismas destacan es la visibilidad regional internacional que se da a su producción a través de LA Referencia y a la interna de la visibilidad que han ganado los repositorios dentro de las instituciones, las mejoras en la calidad de los metadatos y los ajustes o la explicitación de políticas en relación a lo que se publica los repositorios. tener ámbitos colectivos para trabajar sobre problemas y desafíos asociados a la gestión de repositorios, incluso más allá de lo técnico, pensando también en el posicionamiento de los repositorios y de las bibliotecas, que son quienes en general lo gestionan. A nivel personal, el beneficio para mí es contar con un respaldo a todo nivel, primero teórico-conceptual en temas vinculados a políticas de ciencia abierta, la infraestructura de LA Referencia y el continuo soporte técnico que recibimos. Hoy seguimos trabajando en la consolidación de las políticas de acceso abierto a nivel institucional, el soporte para la implementación y consolidación de los repositorios (hay muchas instituciones que recién están desarrollando sus repositorios), la sensibilización de investigadores pensando mecanismos para reportar periódicamente los resultados de cosecha a las instituciones, que es algo que todavía no tenemos resuelto, y empezando a pensar estrategias para fomentar la apertura de datos de investigación, también de la mano de LA Referencia. En conclusión, la participación en LA Referencia ha sido fundamental para un proceso que recién empieza en Uruguay de consolidación de políticas e infraestructura para el acceso abierto.

ANDREA MORA (Costa Rica): Quisiera empezar hablando del modelo en Costa Rica que tiene particularidades completamente diferentes al resto de la región. Por ley en Costa Rica las universidades públicas son las que manejan directamente el presupuesto de investigación, no existe una agencia y entonces

eso permite que se desarrollen iniciativas de formas diferentes para abordar el tema, voy a referirme específicamente a la ciencia abierta. Esto se hace desde un órgano que se llama el Consejo Nacional de Rectores (CONARE), que lo que intenta es articular todas las iniciativas de las cinco universidades. Desde esa perspectiva, las cinco universidades públicas generamos cerca del 80 % de la investigación en Costa Rica, entonces, para poder articular todas las iniciativas que existen vinculadas, específicamente investigación, el CONARE se organiza por subcomisiones y trabajamos para articular las distintas iniciativas de ciencia abierta. Desde el 2014 las cinco universidades públicas, bajo la visión de CONARE, tienen una subcomisión que primero nació como revistas y repositorios, para fortalecer esas dos iniciativas, y se fue transformando en la Subcomisión de Conocimiento Abierto: lo que hace es trabajar en promover todo lo vinculado a la ciencia abierta. En este momento, tenemos dos prioridades: tratar de promover la normativa en acceso abierto, a través de las cinco universidades, y después promover la normativa a nivel nacional. Nosotros también articulamos o tratamos de que todo lo que se genera desde la Subcomisión, como las capacitaciones, queden abiertas a que otras instituciones públicas y privadas se unan a la iniciativa. En nuestros próximos cinco años (2021-2025), nuestros objetivos son: fortalecer los repositorios y fortalecer el cosechador nacional Kimuk; fortalecer las revistas científico-académicas; implementar la propuesta de un sistema CRIS para CONARE; generar espacios de diálogo nacional para la creación de una normativa de conocimiento abierto; promover la cultura de gestión de datos de investigación; genera espacios de gestión del conocimiento en datos abiertos de investigación y consolidar las plataformas tecnológicas necesarias, así como fortalecer las capacidades y acciones en ciencia abierta en la comunidad universitaria. ¿Cómo estamos abordando los desafíos y los beneficios que nos ha traído trabajar con LA Referencia? Desde las cinco universidades nosotros logramos, junto con una delegación que hizo el Ministerio de Ciencia y Tecnología, que se vuelva una realidad poder estar en LA Referencia. ¿Qué hacemos pero abordar todos estos temas que Costa Rica está trabajando? Además de la Subcomisión, nos dividimos en grupos especializados de trabajo

(siete en total) y lo que estamos haciendo es vincular el sector privado a este trabajo para darle seguimiento a esos objetivos que tenemos desde la ciencia abierta. En este momento, tenemos un grupo que evalúa y mejora la tecnología y la parte técnica de los repositorios; un grupo que está trabajando en hacer un nodo de preservación; un grupo que trabaja en hacer un índice de evaluación de revistas científicas; otro grupo de trabajo que está haciendo la normativa CONARE para poder dar impulso a la nacional; otro grupo de trabajo que está trabajando para fortalecer la parte de la gestión de datos de investigación; otro que tiene el nodo de la RDA de Costa Rica y otro grupo que trabaja la visibilidad. Una de las cosas más interesantes de estos grupos de trabajo es que podemos vincular también con las instituciones privadas para darle seguimiento a los distintos objetivos. Nuestra integración a LA Referencia no es sólo desde la parte de repositorios, nosotros hemos logrado en LA Referencia ayuda en la parte de revistas, en la parte de gestión de datos y también incluso esta evaluación de la ciencia abierta, entre otras categorías que trabaja la ciencia abierta, y en conocer experiencias de otros países, cómo lo han hecho y cómo le podemos dar a Costa Rica el poder compartir los conocimientos en la capacitación que nace de esa colaboración, en la tecnología que se utiliza, que se comparte y poder tener acceso a tecnología más actual; incluso participar en tomas de decisiones importantes, no sólo, como decía Paola Azrilevich, en decisiones políticas y técnicas, en tener normas actualizadas, sino en que ya también podemos tener una posición diferente en nuestro país en las redes mundiales de ciencia abierta, en los contactos que hemos ganado también y en los distintos intercambios, creciendo nuestra labor en ciencia abierta. Respecto a los desafíos, hicimos un mapeo de cuáles son los desarrollos que tenemos y uno es el recurso humano especializado en informática, que quisiéramos tener más, y poder incluso colaborar en desarrollos tecnológicos en todas en todas las comunidades que implican la ciencia abierta; nuevas formas de evaluar y alinear la evaluación con respecto a la ciencia abierta pues al igual que Uruguay, no tenemos un marco legal nacional y eso genera muchos problemas. Otro desafío es cambiar la cultura de la investigación a favor de la ciencia abierta, más sensibilización y la capacitación antes de sensibilizarlos, respecto

al acceso abierto; tenemos el desafío de identificar y recuperar los datos de investigación que no sabemos dónde están, si están depositados entre estructuras nacionales y extranjeras; tenemos el desafío y estamos trabajando en cómo se podrían integrar los repositorios de publicaciones con los de datos de investigación, en crear los sistemas CRIS y los ecosistemas de la ciencia abierta y en involucrar, por ejemplo, nuestra comunidad de grupos de trabajo con los de la RDA y, por último, mejorar la calidad de nuestras revistas científicas e intentar que integren prácticas de ciencia abierta.

WASHINGTON SEGUNDO RIBEIRO (Brasil): Desde el IBICT tenemos un papel muy importante, involucrado en el tema de ciencia abierta de Brasil, ya que hacemos el soporte para la creación de repositorios, bibliotecas digitales y revistas electrónicas. También trabajamos en desarrollo de adaptación y difusión de directrices y mantenemos nuestro metabuscador nacional de publicaciones de datos científicos y datos de investigación llamado Oasisbr, que es el nodo de LA Referencia. Principalmente en el tema de la adaptación y difusión de directrices nosotros hemos desarrollado una interfaz de gestión de cumplimiento de las directrices para los gestores de los repositorios. Nosotros le llamamos Studio Dashboard y es realmente fácil: los gestores pueden acceder a su propio perfil y pueden verificar el cumplimiento de las directrices por medio de su propia cuenta, relativo a su repositorio, su revista o su biblioteca digital de tesis y disertaciones. Esto está listo, pero haremos una difusión para todos los miembros de los nodos de LA Referencia y será un desarrollo conjunto con ellos soportado por IBICT. Entonces tenemos un metabuscador que agrega informaciones de las bibliotecas digitales, de las revistas científicas, de los repositorios institucionales de publicaciones científicas y ahora más recientemente los repositorios de datos de investigación y tenemos claro que va a haber por medio de LA Referencia va a llegar OpenAIRE y también Repositorios Científicos de Acceso Abierto de Portugal ([RCAAP](#)) que es una iniciativa similar a Oasisbr pero en contexto portugués y también a la plataforma de LA Referencia. El mismo software permite la exportación a Networked Digital Library of Theses and Dissertations ([NDLTD](#)), que es un consorcio internacional de consorcios nacionales de tesis y

disertaciones. Hoy contamos con más de 600 instituciones que suman más de 1300 fuentes de cosecha y también agregan aproximadamente 2.600.000 documentos. Tenemos una distribución de 110 repositorios institucionales, 127 bibliotecas de tesis y disertaciones locales, 1030 revistas científicas y ocho repositorios de datos ya están agregados a Oasisbr y asimismo RCAAP, nosotros los cosechamos a ellos y viceversa. Ahora voy a hablarles del proyecto BrCris, que también tiene fuerte relación con LA Referencia, que es el ecosistema brasileño de investigación científica. Empieza en 2015 y sólo ahora en 2020 iniciamos su implementación; tiene un equipo muy grande que ha trabajado para su implementación y está muy involucrado en esta última parte del ecosistema de investigación científica, que son los resultados de investigación. En BrCris estamos trabajando todo el ecosistema de investigación y la idea es que la plataforma de LA Referencia esté lista para poder agregar informaciones no sólo de los repositorios sino también de diversas bases de gestión de información científica; principalmente, las bases de perfiles de investigadores brasileños, que es la plataforma LATTES, que tiene más de 60 millones de investigadores y por medio de la plataforma de LA Referencia conseguimos exportar estos datos para ver boards de indicadores de evaluación y también el software VIVO, que es un software para la visualización de grafos que es el tema que estamos tratando acá. Tenemos VIVO ya implementado y ya puede producir visualizaciones. Quería complementar que estamos muy involucrados en construir un ecosistema de investigación brasileña activo, que involucre los repositorios digitales, las revistas científicas, los repositorios de datos de investigación y las plataformas de currículos de los perfiles de investigadores y en obtener también información agregada de bases de gestión información científica, de dentro y de fuera de Brasil, como OpenAIRE, Research Gate y otros. También queremos generar estas interfaces de visualización de la información y recuperación de la información para poder crear indicadores que permitan una evaluación más fiel y más justa en términos de ciencia abierta. Tenemos un grupo también de evaluación para ciencia abierta, que está trabajando fuerte a nivel local en diversas instituciones brasileñas. Las principales dificultades que tenemos son

el mantenimiento de los desarrollos que hacemos, por lo que yo creo que es más fácil construir y más difícil mantener lo que se ha hecho; y también trabajar en distintos temas conexos que son bien interesantes, pero es difícil hacer todo al mismo tiempo. Para concluir, tenemos en Brasil muchas instituciones con mucha diversidad de objetivos y retos, al mismo tiempo que es un problema para nosotros conciliar todas las ambiciones, es también una solución para los problemas que tenemos para congregar el interés nacional.

FREDDY SUMBA (Ecuador): Muy agradecido por estar en esta mesa redonda, con muchas personas que son muy influyentes en la ciencia en América; es muy placentero poder estar con ustedes y yo adelanto las preguntas: para mí es el beneficio porque este grupo sí que consolida las mejores prácticas y las mejores recomendaciones que se puede dar a cada país con respecto a la ciencia abierta como tal. Rápidamente voy a comentar cuáles han sido los principales retos que hemos enfrentado desde CEDIA, que es la Corporación Ecuatoriana para el Desarrollo de la Investigación y la Academia; apoyamos en diferentes aspectos, en ciencia, tecnología e investigación, inversión, a las universidades y sitios de investigación en el país. Se inicia el año 2015 cuando ya establecimos relaciones con LA Referencia, en ese tiempo nos colaboraron con la implementación del nodo nacional de depósito abierto. Cuando hacemos este levantamiento nos encontramos con 22 repositorios, los cuales tenían diferentes malas prácticas que pudimos detectar: pocos opositores, por no hablar de menos de la mitad, contaban con directrices Driver 2.0. Fueron muy precarias las prácticas que se estaban llevando a cabo, a tal punto que en algunos casos ciertas universidades subían documentaciones contractuales, documentos que no tienen como objetivo ser productos de investigación. Lo más importante que logramos en esta evolución es que se incrementaron a 46 repositorios en 2015 y 2016 y en el 2019 tenemos el aval nacional del gobierno para que CEDIA sea el nodo operador de LA Referencia. CEDIA tiene como principio que el sector gubernamental sea representante pero al no tener CEDIA esa etiqueta de alguna forma podríamos decir necesitaban un sistema, lo cual nos permitió tener un mejor acercamiento con las universidades, proveer un mejor soporte y establecer canales de comunicación mucho más efectivos

para lograr la depuración mucho más agresiva de la información y llegar más cerca con las universidades. A tal punto, que desde el 2019 estamos impartiendo capacitaciones con los investigadores y bibliotecarios a nivel nacional; organizamos generalmente webinars y capacitaciones en varias fases. Por ejemplo, en noviembre tenemos un encuentro nacional de investigadores y bibliotecarios en el cual se va a capacitar sobre las nuevas directrices OpenAIRE 4.0. Este cambio desde Driver 2.0 hasta OpenAIRE 3.0 ya implementado en varios repositorios ha sido un reto muy grande y gracias a LA Referencia con la tecnología que tiene el nodo nacional ha sido mucho más fácil de llevar a cabo. Tratamos cada vez de que el transformador de LA Referencia haga menos trabajo y esto ya venga de repositorios normalizados y no se lo cargue al nodo de transformación para que ellos se adapten a los cambios que deberían tener. A costa del 2020 tenemos ya un total de 86 repositorios, cerca del 95% de las universidades del país, con una totalidad de 300.000 recursos hidrográficos cosechados e iniciamos en este punto, con los contactos de LA Referencia que nos han ido guiando en estas prácticas para implementar los primeros prototipos de DSpace CRIS y VIVO. Desde el 2020 hemos evolucionado y ahora estamos trabajando en el observatorio nacional de investigación como punto principal o punto de partida que ha sido el nodo de LA Referencia. La evolución podría decir que ha sido muy fuerte: tenemos ya a las universidades pensando en sistemas CRIS, en perfiles de investigadores y demás, lo cual ha sido un gran reto, tanto a nivel técnico, de capacitación, de adopción, los cuales gracias a LA Referencia los hemos hecho. En cuanto a los beneficios, podríamos decir que nos hemos beneficiado a tal punto que LA Referencia es uno de nuestros puntos principales y centrales dentro del observatorio nacional de investigación y en noviembre lo vamos a comunicar a toda la comunidad científica del país, es parte de nuestra nuestra base principal de información, tenemos cerca de 450 mil recursos bibliográficos actualmente y para Ecuador es una gran cantidad de información para el volumen que tenemos y podríamos decir que es una fuente de información muy importante ya que a través de ella hemos analizado indicadores de decisión, indicadores de calidad, de toma de decisiones a nivel gubernamental y reportería para las

universidades y estamos vinculando varios objetivos secundarios, como por ejemplo, que la Red de Investigadores Semánticos del Ecuador, con el objetivo de que se vinculen varias aristas: se vinculen plataformas de datos científicos, los sistemas CRIS, sistemas de patentes y demás con el objetivo de obtener el ecosistema similar al que comentó Washington hace un momento y en el cual tengamos ya los panoramas que los investigadores necesitan. Esto lo estamos viendo de publicar ahora en noviembre de este año, luego de dos años de desarrollo, y estamos con muchos más prospectos de evolucionar y esperar también que con LA Referencia logremos canalizar esto para que sea un aporte de ciencia. ¿Que tenemos hasta ahora? Que nos hemos beneficiado al tener ya un nodo nacional de LA Referencia de muchos años de técnicas de minería de datos, transformaciones, etc. Ya se ha podido identificar automáticamente perfiles de investigadores y ha logrado tener ya paneles de investigación con datos publicados de diferentes fuentes de información, nodos gráficos de colaboración, índices de colaboración entre investigadores. Hemos logrado por ejemplo detectar cuáles son las tendencias de publicación científica de diferentes tipos de líneas de investigación, donde la gente está investigando, qué está haciendo, qué está publicando, en dónde están publicando; podemos saber qué universidad tiene ciertas filas con áreas de investigación y logramos ya establecer índices de colaboración interinstitucional a través de diferentes tipos de indicadores, como situaciones, referencias, inversión en proyectos colaboración, proyectos de investigación para establecer una margen de colaboración que actualmente no se dispone y esto se lo identificado netamente tratando los datos que están ahora mismo en la organización de LA Referencia. Estos son los grandes beneficios que nos están permitiendo ahora evolucionar ya aún a un nivel de conocimiento a través de la información, a través del tratamiento de estímulos. Estamos trabajando ahora en un proyecto de ciencia de datos, cuyo objetivo es que las universidades dispongan de herramientas de acción a través de técnicas de visualización de información, estamos ya proveyendo de información a las universidades, cuál es su producción, en qué están trabajando, en qué cuartiles, las redes en las que

trabajan y esto es muy importante para que a nivel institucional tomen decisiones.

L. MATAS (Argentina): Quería hacer algunas consideraciones en función de lo que escuchamos y lo que compartieron los compañeros y compañeras de todos los nodos en términos de, por una parte, reflejar muchas de las cosas que se dijeron y resumirlas, pero también de contarles desde LA Referencia cuáles han sido para nosotros los desafíos que hemos enfrentado como red durante estos años de construcción. Más allá de los primeros desafíos, como la consolidación y el pase de un proyecto a una iniciativa regional, la sustentabilidad, el poder tener gobernabilidad y que sea realmente una red federada, que es algo en lo que Alberto Cabezas trabajó incansablemente durante los primeros años, hasta después la parte de lograr tener la presidencia, algo muy importante, que alguien del consejo pueda ser elegido y estar al frente en la representación política frente a otras iniciativas, en este caso y por el segundo periodo Bianca Amaro, más allá de toda esa institucionalización y los acuerdos que hemos logrado la colaboración con Red CLARA y cómo eso ha ido también mejorando con el tiempo y creciendo, todos esos fueron desafíos del principio, pero el desafío constante que nosotros enfrentamos siempre son los cambios políticos en los países, eso es algo que ocurre mucho y que a veces nos genera terremotos a nivel nacional y regional, muchos de ustedes comentaron las interrupciones en los proyectos o el paso de un portal a otro o la discontinuación de la atención. Es algo que nosotros sufrimos y vamos a seguir sufriendo, pero lo bueno de hacer un proyecto político, en el sentido de que tenemos los representantes de los organismos de ciencia y técnica, es que tenemos el poder de la representación en las discusiones, la representatividad de esas instituciones y de ese nivel de gobierno; lo malo de esto es que por supuesto cuando los gobiernos cambian muchas veces los funcionarios cambian, a veces hasta los técnicos cambian y eso nos genera un constante recomenzar y demás. Ahora, la contracara de eso es que el aporte de LA Referencia es también el ser una especie de memoria colectiva y que cuando viene otra gestión no tenga que comenzar de nuevo

porque lo que hacemos con cada uno y muchos de ustedes lo han visto, cuando llega un nuevo consejero porque hubo un cambio, nosotros retomamos el diálogo, le contamos dónde estamos parados, qué habíamos logrado y si cambian los técnicos hay una transferencia tecnológica nueva o una capacitación y eso da continuidad. Nos gusta pensar que eso hace que los países no pierdan o no vuelvan a comenzar, como pasa en nuestra región con muchas políticas de Estado que cuando cambia el gobierno vuelve a empezar o se trata de hacer todo desde cero. Desde LA Referencia, un gran aporte que intentamos hacer es ese, el de preservar la memoria colectiva y hacer que los cambios políticos impacten menos en las políticas de ciencia abierta a nivel nacional y a nivel regional. Muchas veces los recursos de los nodos son un desafío en el sentido de que hay muchas ganas de hacer cosas y hay mucho empuje de parte de los consejeros, pero los recursos son limitados o las infraestructuras, querríamos estar haciendo todos muchas más cosas pero no tenemos como región, y más después del tema del COVID, los presupuestos que queremos tener para dedicar a esto. Creo que ese es otro de los desafíos que enfrentamos siempre como iniciativa: la sustentabilidad tanto de los equipos como también de las propias de los propios instrumentos que tenemos para sostener a LA Referencia.

Ponencias

Eje 1: Comunicación académica,
científica y cultural en abierto

Modelo de gestión de ciencia abierta para la Universidad Autónoma de Zacatecas Hacia la formación de investigadores en prácticas abiertas

MONTSERRAT GARCÍA GUERRERO

Universidad Autónoma de Zacatecas

montsegarcia@uaz.edu.mx

RESUMEN

El presente trabajo expone el modelo de gestión implementado para garantizar el trabajo continuo, alimentación y mejoras de un Repositorio Institucional (RI) en una universidad mexicana, la Universidad Autónoma de Zacatecas (UAZ), lo que derivó en la necesidad de capacitar al colectivo de investigadores sobre las diferentes propuestas de la Ciencia Abierta y las formas de trabajo sugeridas bajo este paradigma, buscando formar a los científicos de la universidad para realizar producción científica y publicaciones de calidad que permitieran el uso del RI de forma sostenida y que culminó con la apertura de la Oficina de Ciencia Abierta. La metodología empleada es un análisis comparativo y exposicional del desarrollo del modelo de trabajo en primera instancia del repositorio y en un segundo tiempo de la Oficina de Ciencia Abierta, creada con el objetivo de ofrecer apoyo integral a los investigadores de la institución, finalizando con la prospectiva de trabajo, así como con la presentación de logros bajo este modelo de acción institucional.

PALABRAS CLAVE

Repositorio institucional; ciencia abierta; oficina de ciencia abierta; apertura de la ciencia; herramientas de investigación.

Introducción

La Ciencia Abierta (CA), como una maduración de prácticas y propuestas de Acceso Abierto (AA), a nivel mundial ha incidido de manera radical en las formas de publicación de la ciencia en todas las áreas; bajo este marco las prácticas de producción y publicación se han visto trastocadas y han dado pie al cambio continuo por parte de todos los involucrados en los procesos científicos actuales (RAMÍREZ MONTOYA, 2015; VICENTE SAEZ *et al.*, 2020). Las propuestas para trabajar bajo este marco siguen sin tener un consenso general pero se pueden encontrar propuestas ya muy avanzadas y respaldadas en lo que se refiere al trabajo de Datos Abiertos, Recursos Educativos Abiertos, Publicación/Literatura Abierta, Investigación Abierta, entre otros, que siguen en la agenda del debate o que ya son parte de la Recomendación UNESCO de Ciencia Abierta (2021). La mayoría de estos elementos tiene como eje el uso de un repositorio donde poder colgar los archivos o documentos de acuerdo a las diversas necesidades de las propuestas o ejes de la taxonomía de CA (OPEN SCIENCE TRAINING INITIATIVE, 2018).

La necesidad de la promoción y apertura de la ciencia ha sido un tema de discusión recurrente desde que MERTON (1977) habló del comunismo de la ciencia, como la necesidad del conocimiento como un elemento abierto entre la comunidad científica, y desde entonces la búsqueda por lograr compartir los resultados de los esfuerzos científicos ha sido un elemento común en las instituciones de educación e investigación del mundo. Esta necesidad impulsada por la legislación venía acompañada de formas de trabajo académico cada vez más influenciadas por corrientes de apertura de todo el ciclo de trabajo científico, corrientes que llegaron cada vez a más áreas de la ciencia y que redundaron en una visión dominante de necesidad de acceso a los resultados científicos y a un buen número de prácticas encaminadas en este sentido, adoptadas de forma cada vez más extensiva.

Es a inicios de los años dos mil que en las agendas de los gobiernos e instituciones cobra importancia el tema del AA y la CA, con diversas propuestas de acción centradas en el uso y manejo de la información; para BABINI y ROVELLI.

la Ciencia Abierta busca «alternativas que mejoren los procesos de investigación individual y colaborativa, su comunicación y reproducibilidad a fin de acelerar la producción y uso de nuevos conocimientos en la sociedad» (2020: 40). Todo esto incluye también la perspectiva del derecho humano a la ciencia, buscando construir experiencias más inclusivas alrededor de la ciencia y los repositorios han sido parte nodal de todas las propuestas de acción: «los repositorios institucionales, que es la vía prioritaria de implementación de las políticas de acceso abierto a nivel institucional y la modalidad apoyada por las leyes nacionales de acceso abierto aprobadas en la región» (BABINI & ROVELLI, 2020: 64). Los conceptos o temas básicos dentro de este esquema son que el conocimiento debe ser *encontrable, accesible, interoperable y reusable*.

Si bien las exigencias y las prácticas, en el sentido de la apertura de la ciencia, se iniciaron en Europa con propuestas de legislación y normas de trabajo que venían respaldadas por la Unión Europea, poco a poco los demás países se fueron alineando a esta forma de trabajo y los gobiernos realizando trabajo legislativo en ese sentido. La Universidad Autónoma de Zacatecas no tuvo, en cincuenta años de historia, ningún sistema de compilación de la producción científica de la institución; los indicadores en este sentido eran inexistentes y solo se encontraban esfuerzos aislados de informes por unidad académica. Un elemento decisivo para iniciar con el esfuerzo del Repositorio Institucional, como primer proyecto para compilar y hacer accesible la producción científica, fue la legislación mexicana que venía empujada por procesos mundiales que tienen como antecedentes las declaraciones de Budapest, Berlín y Bethesda y que culminan en la reforma a la Ley de Ciencia y Tecnología en México en el año 2014, reforma con la que se norma la implementación de un Repositorio Nacional y el apoyo para la creación de Repositorios Institucionales; unos años después surge en la región latinoamericana la Declaración de Panamá sobre Ciencia Abierta en el 2018.

Puesta en marcha del Repositorio Institucional de la UAZ

Buscando responder a las tendencias de AA y CA, la propuesta de acercamiento inicial de la Universidad Autónoma de Zacatecas (UAZ) fue la implementación de un Repositorio Institucional (RI) en el año 2016, que trabajara por medio de encargados por departamento o área de la institución (Unidades Académicas). El trabajo inició como parte de un proyecto apoyado por el Consejo Nacional de Ciencia y Tecnología (CONACYT), con un periodo de duración de dos años (2016-2017), periodo en que se trabajó con una sola unidad académica y donde se plantearon elementos iniciales como el nombre, la imagen y las gestiones institucionales para lograr no solo el apoyo por parte de la administración central, sino la confianza de los investigadores para depositar sus trabajos.

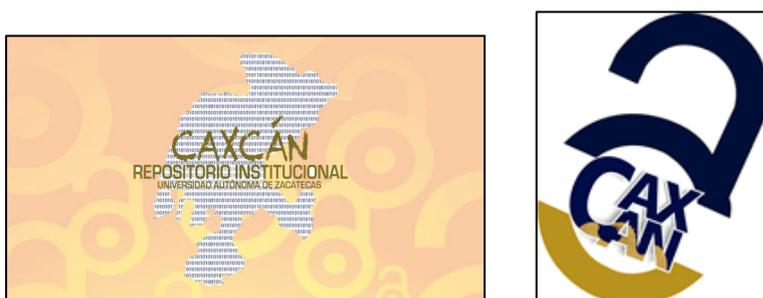


FIGURA 1. LOGOS RI-UAZ 2017 Y 2021

La FIGURA 1 da cuenta del desarrollo que tuvo el logo del RI de la UAZ, que evolucionó drásticamente en el período, para tener una imagen más apegada a la imagen institucional con apoyo del equipo de imagen y comunicación social de la universidad. Lo mismo sucedió con la imagen de la plataforma en general, que tuvo una imagen limitada al inicio, usando la base de Dspace, y evolucionó hacía una versión más trabajada, como se puede observar en la FIGURA 2. Hubo mejoras en la programación de plantillas y la vista se cambió de XMLUI a JSPUI, así como la vista de la página fue mejorada de forma que la experiencia de visita fuera más accesible y agradable para el usuario.

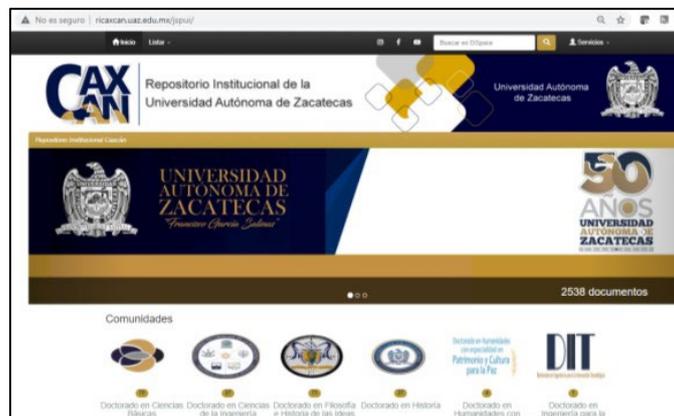
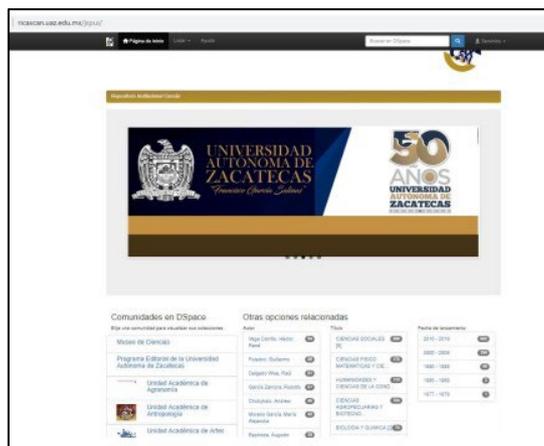


FIGURA 2. REPOSITORIO UAZ 2017, 2019 Y 2021

La forma de trabajo del RI también evolucionó, pues al proceder un tiempo bajo el esquema de responsables se vio la imposibilidad de seguir trabajando de esta manera, pues aquellos encargados habían sido designados con mayor carga de trabajo sin obtener pago por la tarea extra, lo que derivó en la

necesidad de trabajar bajo una propuesta de auto-depósito, lo que a la larga también permitiría que el RI se siguiera manteniendo alimentado sin importar quienes trabajen o se encuentren al frente del proyecto.

Para realizar la propuesta de trabajo por auto-depósito por parte de los investigadores, se emprendió un proceso de capacitaciones uno a uno, lo que derivó en un mapeo de las necesidades de información del colectivo académico de la UAZ, que se resumen de la siguiente manera: identidad digital del investigador, acercamiento a circuitos científicos de las diferentes áreas del conocimiento, derechos de autor, cómo reconocer editoriales depredadoras, cómo seleccionar una revista para publicar, buenas prácticas en la publicación científica, uso de gestores académicos, redes sociales académicas, uso de gestores bibliográficos, y escritura académica. Con esta información se procedió a promover un catálogo de cursos cortos que respondieran a las necesidades encontradas en la universidad.

Para promover la iniciativa de CA a toda la universidad y, debido a la importancia de la difusión del conocimiento y de la actividad universitaria a nivel local, nacional e internacional, en el 2018 el Repositorio Caxcán dejó de ser parte de una Unidad Académica para convertirse en una plataforma Institucional gracias al apoyo y acogimiento de la Coordinación de Investigación y Posgrado. En ese momento, se contaba con una comunidad y la participación de veinte investigadores; gracias al apoyo financiero de CONACYT se tenía un hosting externo que garantizaba el cumplimiento del compromiso que se estableció en ese proyecto de mantener su funcionamiento 24 horas al día los 7 días de la semana.

A partir del año 2019 y con apoyo de un segundo proyecto CONACYT, el Repositorio Institucional Caxcán se encuentra alojado en un hosting interno institucional, que permite la preservación de la plataforma a largo plazo. Se logró expandir el conocimiento y uso acerca del RI a través de la concientización de la apropiación de esta plataforma, como medio de potenciar no sólo la visibilidad personal sino de toda la institución y del país, con una capacitación intensiva a profesores de diversas unidades académicas, con un

alcance de 32 unidades académicas y alrededor de 600 profesores que han ingresado elementos de producción académica. Algunas de las actividades principales para poblar y dar a conocer el RI fueron:

- Apoyo en carga de historial de documentos académicos a profesores.
- Escaneo de material histórico para inclusión en el RI, para el aumento de la preservación digital.
- Ingreso del RI al Directory of Open Access Repositories (DOAR) y al Registry of Open Access Repositories (ROAR), así como en la Red Mexicana de Repositorios Institucionales (REMERI).
- Seguimiento de requisitos CONACYT para la continua cosecha de ítems.
- Gestión para inclusión de tesis de posgrado, sobre todo las de programas del Padrón Nacional de Posgrados de Calidad (PNPC).

Derivado de los esfuerzos anteriores y de difusión de la plataforma y el acercamiento a las prácticas abiertas, los principales logros del RI a la fecha son:

- Para el año 2021, el RI cuenta con alrededor de 2.000 ítems cosechados por el Repositorio Nacional y más de 2.500 elementos susceptibles de ser consultados a texto completo.
- Ampliación y creación, a través del desarrollo y expansión del RI, de nuevas redes y alianzas académicas que permitan crecer como universidad.
- Posicionamiento en el puesto 7 del RN de CONACYT de acuerdo al número de ítems y las consultas de la plataforma.

Obligación y necesidad de prácticas abiertas

El trabajo para la puesta en marcha del RI incluyó un proceso de concientización de por lo menos dos años, pues no existía ninguna experiencia similar en la universidad; de ahí surge la necesidad de la Oficina de Ciencia

Abierta (OCA) y de las propuestas de trabajo para esta oficina; de esta forma el acercamiento ha sido de forma consistente, integrando cada vez un mayor número de elementos de apoyo y de herramientas útiles para atender la realidad y necesidades actuales de los investigadores para la publicación de sus trabajos.

Un elemento que dio impulso a la propuesta del uso del RI de manera generalizada fue la inclusión, dentro de los parámetros del Padrón Nacional de Posgrados de Calidad (PNPC) del Consejo Nacional de Ciencia y Tecnología (CONACYT), de la exigencia de depositar todas las tesis de los posgrados reconocidos como de calidad en un RI que fuera cosechado por el Repositorio Nacional (RN) en el año 2020, lo que dio pie al depósito en el RI del trabajo de tesis final que da acceso al grado académico. Este agregado a las exigencias para pertenecer o consolidarse en el PNPC funcionó en un doble sentido: en primera instancia para acercar a los estudiantes de posgrado a estas prácticas y en segunda acercar a los asesores al depósito no solo de las tesis sino de la producción en el RI y por tanto a las prácticas abiertas.

Los acercamientos realizados, así como las exigencias de los diferentes programas, han permitido que a la fecha se hayan capacitado a más de 600 profesores de los 1.000 mapeados al inicio de la puesta en marcha del RI, cuidando que se tuviera acercamiento con todas las unidades académicas (34), de las 8 áreas en que está dividida la universidad, para con esto tener presencia en el RI de todas las áreas del conocimiento que se trabajan en la institución. El énfasis de la atención se puso en los docentes/investigadores con carga laboral en programas de posgrados, con 35 programas de nivel maestría y 14 de nivel doctorado (con menos de la mitad de ellos reconocidos por el PNPC) y en propuestas de acercamiento para este nivel. Poco a poco se fue trabajando con la mayoría de los investigadores con reconocimientos por su producción académica: para el caso de México ser parte del Sistema Nacional de Investigadores (SIN) y el reconocimiento como parte del Programa de Mejoramiento del Profesorado (PRODEP).

Un elemento que se obtuvo del acercamiento con los investigadores universitarios fue la necesidad de capacitar en prácticas abiertas de investigación (BOWMAN & KEENE, 2018) y en derechos de autor, pues las tendencias de publicación en áreas básicas han promovido desde hace décadas el envío de trabajos a grandes editoriales, reconocidas como prestigiosas que fueron cambiando sus modelos económicos a partir de la CA (GARCÍA & FERNÁNDEZ GALÁN, 2021); por otro lado, los investigadores de las áreas de Ciencias Sociales y Humanidades tiene un acercamiento incipiente en estas prácticas de publicación como lo demuestra la cantidad de revistas indexadas que publican trabajos dentro de estas áreas del conocimiento.

De repositorio de literatura a mixto

El RI institucional nació como una propuesta de depósito de documentos académicos mayormente, con la inclusión de pocas tesis de posgrado; para el año 2020, con la adición en el PNPC del requisito de tener las tesis en un RI el número de solicitudes de inclusión de tesis se incrementó de forma importante, pasando de albergar 26 tesis en 2017 y poco más de 40 tesis a inicios del 2020 a cerca de 300 para julio de 2021 (TABLA 1). La forma de depósito de las tesis se hace por medio del Responsable de Programa de PNPC o una persona designada por el mismo, ya que trabajar la inclusión de tesis por auto-depósito requeriría la alta de un mayor número de usuarios en el repositorio, que difícilmente seguirán utilizando la plataforma y que representan una necesidad de memoria más robusta, por lo tanto la carga se ha definido de forma institucional.

TIPO	2017	2019	2021	Gran total
Total de elementos incluidos	284	604	1.659	2.547
Multimedia	0	0	6	6
Tesis	26	46	289	361
Datos	0	0	1	1
REA	0	0	1	1

TABLA 1. CRECIMIENTO DE RI UAZ 2017-2021

Elaboración propia con información de los proyectos CONACYT y del RI

La TABLA 1 muestra, además de la información sobre las tesis integradas, el desarrollo del RI de la UAZ en el período 2017-2021, en el que se pasa de 284 elementos totales integrados en el informe a CONACYT en el año 2017, alcanzando casi 900 para el informe CONACYT del año 2019, a más de 2.500 elementos en julio de 2021, con un incremento de más del mil por ciento en un período de 4 años. Esto da muestra de cómo la plataforma se ha vuelto una herramienta primordial en el desarrollo de la producción científica de la institución y como cada día se incluyen mayor número de elementos sin ser necesidad de un proyecto específico, sino como el desarrollo natural de la institución, cuyos investigadores han incluido entre sus prácticas de visibilidad de producción la carga en el RI.

Para el año 2020 también se recibió la primera solicitud para la inclusión de archivos de datos de investigación, como resultado del requisito por parte de una publicación internacional donde se buscaba hacer una publicación del área de Ciencias de la Salud; esto hizo que para el segundo semestre del año se incluyera una nueva colección de Datos, que por ahora aparece sólo en la comunidad de la Unidad Académica de Medicina Humana y Ciencias de la Salud. El formulario se encuentra disponible para cualquier otra comunidad que decida integrar esta colección entre sus opciones de carga y representa ya la oportunidad de manejar este tipo de archivos en la plataforma.

Otro logro más reciente es el trabajo que se está realizando, como parte de un proyecto apoyado por el CONACYT, para realizar la sistematización de talleres de ciencia recreativa como Recursos Educativos Abiertos (REA) y que puedan ser replicados de forma abierta a nivel amplio, que derivará en el Repositorio Nacional de Ciencia Recreativa albergado en el RI de la UAZ. Los REA son parte importante del ecosistema de CA, como una propuesta de apertura de experiencias educativas que pueden ser aprovechadas en distintos momentos y contextos (BUCAREY, 2017; HILTON, 2020; RIVERA *et al.*, 2011). El uso de este tipo de formatos, para poner en AA modelos de ciencia recreativa, resulta una poderosa herramienta para facilitar el avance a grupos nuevos, enriquecer el trabajo de equipos con más experiencia y ofrecer propuestas valiosas para profesores de diferentes niveles educativos.

De un Repositorio Institucional a una Oficina de Ciencia Abierta

El conocimiento que tenían los investigadores de las prácticas científicas actuales daba cuenta de la forma en que se estaban formando los estudiantes en los diferentes niveles, sobre todo posgrado y licenciatura, lo que derivaba en una formación con un sesgo sobre los cambios de los últimos años en lo que se refiere a la forma de producción y publicación de resultados científicos y académicos, cambios fuertemente influenciados por la corriente de CA, que han derivado en nuevas formas de acceder a la publicación científica con la ruta verde, ruta dorada, ruta híbrida y ruta diamante (UNIVERSO ABIERTO, 2019). La preocupación más urgente era para el nivel posgrado, donde se supone el estudiante debe estar aprendiendo a investigar y la formación debería incluir las prácticas de publicación científica, así como los circuitos y el acercamiento y acompañamiento en los procesos de publicación.

Con esta inquietud, se procedió a la oferta de 10 cursos hacia la comunidad docente, brindándolos desde el año 2017 en las Escuelas de Invierno y Verano promovidas por las autoridades de la UAZ y cada vez más bajo una agenda de

trabajo por medio de solicitudes específicas de las Unidades Académicas que buscaban formar a su planta docente para la incentivación de la producción académica. Como resultado del aumento en la carga de trabajo y la importancia de las actividades para el año 2018 se promovió la creación de la Oficina de Ciencia Abierta, propuesta que fue apoyada para las autoridades universitarias en ese mismo año y que trabaja bajo seis ejes de acción continua:



FIGURA 3. EJES DE TRABAJO DE LA OFICINA DE CIENCIA ABIERTA, UAZ

Elaboración propia

La FIGURA 3 ilustra el hecho de que el RI de la institución sirve como eje articulador de las otras propuestas de trabajo para la apertura de la ciencia de la institución, destacando el hecho de que las otras propuestas y la propia oficina surgen a partir de la propuesta de trabajar un repositorio para la UAZ. La Oficina de Ciencia Abierta es resultado del proyecto de repositorio y los otros ejes han surgido como necesidad para la implementación y buen uso y manejo de la plataforma y de esta forma es como se ha logrado una propuesta integral

de fomento y apoyo para la apertura de la ciencia por parte de los investigadores. Lo que está en consonancia con las tendencias internacionales donde los RI fungen como eje de las acciones de apertura de la investigación en todos sus niveles.

Actualmente, la plataforma de revistas UAZ¹ se maneja bajo el modelo de Open Journal System (OJS) del Public Knowledge Project (PKP) y cuenta con quince revistas activas; siete de ellas tienen un desarrollo histórico y ocho son de nueva creación. Todas ellas tienen equipos editoriales que han sido capacitados por personal de la OCA en la implementación de los procesos y las buenas prácticas editoriales. Adicionalmente, las ya existentes recibieron apoyo en migrar su información histórica a la plataforma de OJS y en buscar la indexación nacional e internacional. Esta línea de trabajo ha logrado la indexación y aplicación de buenas prácticas editoriales de las revistas de la universidad y la creación de nuevas propuestas de calidad (pasando de siete a quince revistas del 2018 al 2020), la migración del modelo impreso al digital, la confianza de editores y autores en la OCA, así como un proceso de difusión en redes sociales, radio y televisión para el apoyo a la visibilidad de los esfuerzos de publicaciones periódicas de la institución.

Como parte de los esfuerzos por promover la visibilidad de los esfuerzos científicos y académicos de la institución se ha iniciado un proceso de apoyo en diferentes gestiones administrativas por parte de la OCA. Centrándose principalmente en acompañamiento para solicitud de ingreso al PNPIC y al Sistema Nacional de Investigadores (SNI), que ha culminado en el ingreso de más de veinte investigadores en el año 2020 y en un proceso de apoyo a más de treinta investigadores en el año 2021; referente al PNPIC, para el 2020 se apoyaron once programas, siendo reconocidos todos dentro del padrón y para el año 2021 se apoyan a todos los programas de nuevo ingreso y de reevaluación.

Sobre la garantía de que las publicaciones y tesis de posgrado sean productos originales y de calidad, en marzo del año 2020 se inició la gestión

¹ Sitio web: www.revistas.uaz.edu.mx

para la contratación del servicio de software de detección de similitudes, que maneja la OCA desde su adquisición y que ha permitido cumplir con los requisitos para el PNPC desde entonces, y que permite identificar las similitudes y mejorar los trabajos de grado y de investigación de alumnos y docentes de la institución. El servicio con historia de un año en la institución es cada vez más utilizado en primera instancia por los posgrados PNPC y cada vez más por investigadores de las diferentes áreas del conocimiento.

En lo referente a la oferta de cursos y capacitación, a la fecha se ha atendido un universo de alrededor de 600 investigadores de todas las áreas de la universidad, en temas como identidad digital, redacción de artículos científicos, derechos de autor, buenas prácticas e investigación, entre otros. Se ha capacitado a más de 300 alumnos en el uso de fuentes de información y bases de datos abiertas.

Además, se ha ampliado el alcance de la OCA fuera de la universidad dando capacitaciones en el Hospital General de Zacatecas y en el Centro de Actualización del Magisterio del Estado, así como a diferentes editoriales que dan servicios a investigadores de la universidad.

Por último, se ha implementado un programa de difusión no solo de las actividades de la OCA sino de los esfuerzos científicos y académicos de la universidad, que ha tenido entre las estrategias la implementación de una página de Facebook, que tiene actualmente alrededor de 2.500 seguidores, la participación en un programa de TV del canal de Sistema Zacatecano de Radio y Televisión, así como participación en el programa radiofónico de la universidad donde se invita a responsables de revistas y a investigadores con proyectos de frontera a presentar sus trabajos, resultados y convocatorias. Además, se ha buscado la participación del equipo de trabajo de la OCA en eventos nacionales e internacionales, con especial enfoque en los promovidos por la UAZ, para la promoción de la consulta y uso del RI de forma extensiva.

Todas las estrategias reseñadas previamente han permitido que el uso del RI no sea solamente para la carga de la producción de los investigadores sino que la plataforma se use para la consulta de los estudiantes e investigadores. El

hecho de que los seis ejes se encuentren entrelazados ha permitido un desarrollo homogéneo de las diferentes aristas de acción que han redundando en el incremento de indicadores de la universidad y en la promoción de prácticas abiertas dentro de la población docente y estudiantil.

Prospectiva para la promoción del AA en la institución

Los seis ejes bajo los que trabaja la OCA representan la respuesta institucional para brindar herramientas de trabajo y apoyo a las iniciativas de apertura de la ciencia que cada día son más importantes en la vida académica alrededor del mundo, pues los diferentes evaluadores de calidad, que para México son el SNI, el PNPC, el PRODEP, entre otros, cada día promueven más prácticas abiertas y opciones de apertura para la ciencia; de esta forma, resulta necesario que la institución cuente con los elementos de apoyo en este sentido. Los cambios se suceden de forma rápida y es por eso que la OCA actualmente se encuentra en proceso de evaluación y adecuación para trabajar un sistema de información institucional integral, y la propuesta es migrar hacia DSpace CRIS, que permita compilar en la misma plataforma no solo la producción sino la información de cada investigador de la institución.

Algunos de los temas que se encuentran en desarrollo actualmente y que son necesidades a corto plazo son la inclusión de RI-Caxcán en las políticas universitarias como eje de acción primordial junto a los otros elementos considerados por la OCA, como una campaña de emisión de ORCID para la migración como identificador de los investigadores, de forma que se pase a un proceso de internacionalización de la identidad digital, la adquisición de DOI para elementos del RI y la plataforma de revistas como un proceso de identificación persistente de la producción UAZ.

A mediano plazo se busca gestionar un ecosistema tecnológico de CA (GARCÍA HOLGADO & GARCÍA PEÑALVO, 2020; GARCÍA PEÑALVO, 2017; WELLER & VAN LOENEN, 2017) para la UAZ con el objetivo de crear un programa de vinculación y divulgación con todos los que generan ciencia en la institución: datos, libros, revistas y otros para fomentar la visibilidad. Lo anterior, con la firme convicción

de que la UAZ tiene grandes recursos académicos para ser parte de la corriente principal en muchas áreas de la ciencia y por eso es necesario trabajar en la promoción de un proceso de visibilidad y calidad constante, que permita posicionar a la UAZ dentro de las principales instituciones del país y la región.

A largo plazo se pretende que la UAZ sea parte de proyectos de visibilidad de la producción científica de la región, lo que incluye métricas alternativas y otras propuestas de medición de la ciencia que permitan acortar las asimetrías existentes para la visibilidad de la producción de países en desarrollo en relación con los desarrollados, propuestas que serán posibles por la relación con responsables de RI de la región y por seguir trabajando en propuestas abarcadoras en este sentido, donde se comparte la experiencia de la UAZ.

Conclusiones

El tema del AA y la CA es una preocupación que ya tiene varias décadas y que se presume de largo aliento, pues la agendas de diferentes organismos internacionales, países e instituciones han estado enfocadas en estos temas desde hace años, lo que hace necesario que los investigadores y las instituciones tengan respuestas efectivas a todo el movimiento que va ofreciendo líneas de acción cada vez más especializadas.

Las acciones reseñadas en este texto dan cuenta de cinco años de trabajo, bajo un proceso que empieza con la idea de implementar un RI en la institución y que culminó con lo que se considera un ecosistema social y tecnológico de apoyo, que ha permitido adquirir la confianza de la comunidad universitaria y un trabajo respetuoso con todas las unidades académicas y áreas, promoviendo que la universidad brinde las herramientas necesarias a los investigadores que quieran ser acreedores a reconocimientos, proyectos de investigación o que se quieran acercar a la publicación científica en abierto. A partir del trabajo de la OCA, la UAZ tiene una plataforma donde se puede mostrar, de manera totalmente abierta, la producción científica y otra en donde se puede promover la publicación de acuerdo a los más altos estándares de calidad, logrando la inclusión de la UAZ en la corriente de apertura de la ciencia que viene

cambiando los patrones de acción de la comunidad científica en todo el mundo, con esfuerzos por realizar una «ampliación de plataformas en las que investigadoras e investigadores comparten datos, publicaciones, experimentos y equipamiento» (BABINI & ROVELLI, 2020: 161).

Se busca seguir trabajando para posicionar a la UAZ en los mejores indicadores de forma que sea reconocida también por su carácter científico, académico e innovador, y que esto impacte también en los indicadores y la forma de producción de la región. Por tanto, la idea de un ecosistema tecnológico de CA más robusto pretende ser el paso a seguir, una vez que se tiene el apoyo no solo de la Secretaría Académica, la coordinación de investigación y posgrado, sino de la oficina de información institucional y de los encargados de sistemas tecnológicos de la institución, con quienes se pretender colaborar para la gestión y puesta en marcha de una propuesta integral de información institucional en abierto que permita dar accesibilidad a la información de la universidad.

Referencias

- BABINI, D. y ROVELLI, L. (2020). *Tendencias recientes en las políticas científicas de ciencia abierta y acceso abierto en Iberoamérica*. CLACSO; Fundación Carolina.
- BOWMAN, N. D. y KEENE, J. R. (2018). A Layered Framework for Considering Open Science Practices. *Communication Research Reports*, 35(4), 363-372. <https://doi.org/10.1080/08824096.2018.1513273>
- BUCAREY, S. G. y AGUILAR, M. L. (2017). Recursos Educativos Abiertos en la Facultad de Medicina de la Universidad Austral de Chile, proyecto AUS1410. *Formación Universitaria*, 10(2), 23-30. <https://dx.doi.org/10.4067/S0718-50062017000200004>
- GARCÍA HOLGADO, A., GARCÍA PEÑALVO, F. J. y BUTLER, P. (2020). Technological ecosystems in citizen Science: A framework to involve children and young people. *Sustainability*, 12(5), 1863. <https://doi.org/10.3390/su12051863>
- GARCÍA PEÑALVO, F. J. (2018). Ecosistemas tecnológicos universitarios en J. Gómez (Ed.), *UNIVERSITIC 2017. Análisis de las TIC en las Universidades Españolas* (pp. 164-170). Crue Universidades Españolas.

- GARCÍA, M. y FERNÁNDEZ GALÁN, C. (2021). Modelos de evaluación de la Educación superior: los circuitos de comunicación científica en México. *Yeiya* 2(1), 5-16. <https://doi.org/10.33182/y.v2i1.1549>
- HILTON III, J. (2020) Open educational resources, student efficacy, and user perceptions: a synthesis of research published between 2015 and 2018, *Educational Technology Research and Development*, 68, 853-876. <https://doi.org/10.1007/s11423-019-09700-4>
- MERLO VEGA, J. A. (Ed). (2018). *Ecosistemas del Acceso Abierto*. Universidad de Salamanca.
- MERTON, R. (1977). *La sociología de la ciencia*. Alianza.
- OPEN SCIENCE TRAINING INITIATIVE. (2018). *Open Science Training*. <http://www.opensciencetraining.com/content.php>
- RAMÍREZ MONTOYA, M. S. (2015). Acceso abierto y su repercusión en la Sociedad del Conocimiento: Reflexiones de casos prácticos en Latinoamérica. *Education In The Knowledge Society (EKS)*, 16(1), 103-118. <https://doi.org/10.14201/eks2015161103118>
- RIVERA MORALES, R., LÓPEZ IBARRA, A. Y RAMÍREZ MONTOYA, M. S. (2011). Estrategias de comunicación para el descubrimiento y uso de Recursos Educativos Abiertos. *Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación (REICE)*, 9(4), 141-157.
- SCHILTZ, M. (2018). Science Without Publication Paywalls: cOAlition S for the Realisation of Full and Immediate Open Access. *PLoS Medicine* 15(9), e1002663. <https://doi.org/10.1371/journal.pmed.1002663>
- SCIENCE EUROPE (2018). cOAlition S. Accelerating the transition to full and immediate Open Access to scientific publications. https://www.coalition-s.org/wp-content/uploads/PlanS_Principles_and_Implementation_310519.pdf
- UNESCO (2019). Hacia una recomendación de la UNESCO sobre la ciencia abierta. https://en.unesco.org/sites/default/files/open_science_brochure_sp.pdf
- UNESCO (2021). Proyecto de recomendación de la UNESCO sobre la ciencia abierta. https://unesdoc.unesco.org/ark:/48223/pf0000376893_spa?posInSet=1&queryId=N-EXPLORE-8094371a-c1b3-4727-81bb-ba237f38409e
- UNIVERSO ABIERTO (2019). Las 5 rutas para llegar al acceso abierto: verde, dorada, bronce, híbrida y diamante. <https://universoabierto.org/2019/03/13/las-5-rutas-para-llegar-al-acceso-abierto-verde-dorada-bronce-hibrida-y-diamante/comment-page-1/>
- VICENTE SAEZ, R. y MARTÍNEZ FUENTES, C. (2018). Open Science now: A systematic literature review for an integrated definition. *Journal of Business Research*, 88, 428-436. <https://doi.org/10.1016/j.jbusres.2017.12.043>

VICENTE SAEZ, R., GUSTAFSSON, R., y VAN DEN BRANDE, L. (2020). The dawn of an open exploration era: Emergent principles and practices of open science and innovation of university research teams in a digital world. *Technological Forecasting and Social Change*, 156. <https://doi.org/10.1016/j.techfore.2020.120037>

WELLER DONKER, F. y VAN LOENEN, B. (2017). How to assess the success of the open data ecosystem? *International Journal of Digital Earth*, 10(3), 284-306. <https://doi.org/10.1080/17538947.2016.1224938>

Lineamientos institucionales sobre Ciencia Abierta: caso de la Universidad del Rosario en Colombia

MARÍA LUCÍA LIZARAZO RIVERO

Universidad del Rosario

marial.lizarazo@urosario.edu.co

MALGORZATA LISOWSKA NAVARRO

Universidad del Rosario

margarita.lisowska@urosario.edu.co

HUMBERTO BLANCO CASTILLO

Universidad del Rosario

humberto.blanco@urosario.edu.co

RESUMEN

En las últimas décadas, el movimiento de Ciencia Abierta ha cobrado un papel protagónico en los procesos de generación de conocimiento a nivel mundial. Su implementación, impulsada entre otras cosas por los avances del mundo digital, aborda un cambio en diversos aspectos del proceso científico dado que «pretende hacer la ciencia más abierta, accesible, eficiente, democrática y transparente» (UNESCO, 2020). Teniendo en cuenta su enfoque disruptivo e interdisciplinar, la transición hacia la ciencia abierta representa un reto para las instituciones de educación superior, pues exige un cambio en la cultura y los procesos orientado hacia una mayor conexión con las necesidades de la sociedad, y con los diferentes actores de la comunidad que lo rodea; sin embargo, aún falta una comprensión global de lo que significa, así como de las oportunidades que ofrece y los retos que plantea (UNESCO, 2020). Este trabajo describe la

experiencia de la Universidad del Rosario en la construcción de una estrategia institucional para la materialización de la Ciencia Abierta y tiene como objetivo contribuir a la reflexión regional alrededor del tema para dar a conocer los desafíos y motivar a otras instituciones a emprender este camino.

PALABRAS CLAVE

Ciencia abierta; lineamientos institucionales; política institucional; modelo institucional; Open Science.

Introducción

Ciencia Abierta, en su sentido más amplio, se define como «el conjunto de esfuerzos orientados a lograr que el proceso científico sea más abierto e inclusivo a todos los actores relevantes, dentro y fuera de la comunidad científica» (OCDE, 2017). Su materialización representa un gran reto, porque requiere de una reflexión institucional acerca de la forma en que se crean, almacenan, comparten y difunden todos los productos asociados al desarrollo de investigación.

La Universidad del Rosario ha asumido este reto, y hace más de una década ha desarrollado una estrategia enfocada en la promoción y materialización del acceso abierto, haciendo parte y liderando iniciativas como:

- La participación activa en iniciativas como la Biblioteca Digital Colombiana², LA Referencia³ y BIREDIAL-ISTEC⁴.

² La [Biblioteca Digital Colombiana \(BDCOL\)](#) es la Red Colombiana de Repositorios y Bibliotecas Digitales que indexa toda la producción académica, científica, cultural y social de las instituciones de educación superior, centros de investigación, centros de documentación y bibliotecas en general del país.

³ [LA Referencia](#) o Red Federada de Repositorios Institucionales de Publicaciones Científicas es una red latinoamericana de repositorios de acceso abierto. Por medio de sus servicios, apoya las estrategias nacionales de Acceso Abierto en América Latina mediante una plataforma con estándares de interoperabilidad, compartiendo y dando visibilidad a la

- La creación, en el 2008, del [Repositorio Institucional EdocUR](#), el cual consolida y brinda acceso a los documentos producidos por la universidad en su función docente, investigativa y de extensión.
- La implementación, en el 2010, del [Portal de Revistas Científicas](#), a través del cual se difunden en acceso abierto todas las revistas académicas y de divulgación, editadas por la universidad, sin generar ningún costo para los autores.
- La expedición, en el 2017, de la [Política Institucional de Acceso Abierto](#), siendo así la primera universidad en Colombia en contar con una política de este tipo, que aborda los lineamientos para materializar el acceso abierto de la producción académica de la universidad.
- Firma de la [Declaración de Berlín](#) sobre Acceso Abierto al Conocimiento en Ciencias y Humanidades, como la primera universidad colombiana, sumándonos a más de 450 instituciones en el mundo.
- Y la generación, desde 2018, de apoyo financiero para el pago de costos de artículos de investigación que sean publicados en abierto. Este apoyo se canaliza a través de la Dirección de Investigación e Innovación y está destinado a la publicación en las revistas de acceso abierto ubicadas en los mejores cuartiles de sus disciplinas. Adicionalmente, se ofrece un servicio de acompañamiento a los investigadores, para seleccionar la revista en donde se va a publicar. Ambas estrategias garantizan la pertinencia de las revistas y la destinación de la financiación en pro de la calidad.

producción científica generada en las instituciones de educación superior y de investigación científica.

⁴ [BIREDIAL-ISTEC](#) es una conferencia anual dedicada a discutir, analizar y promover las iniciativas institucionales, nacionales y regionales relacionadas con el acceso abierto al conocimiento y la visibilidad académica latinoamericana. El evento es un escenario propicio para compartir el conocimiento, conocer los avances y establecer contactos regionales sobre el rol del acceso abierto, la gestión de los datos de investigación y los datos públicos; la gestión de repositorios institucionales y temáticos; los derechos de autor y la reproducibilidad de la investigación.

De igual manera, la Universidad ha propiciado el desarrollo de otros componentes de la Ciencia Abierta, como:

- Sistema de Integridad Científica: en el año 2017 se ha consolidado el sistema, que acompaña a los investigadores en el cumplimiento de las buenas prácticas en investigación.
- Datos de investigación: la expedición, en el 2019, de la [Política institucional de Gestión de Datos de Investigación](#) y la implementación del [Repositorio de Datos de Investigación](#).

Estos esfuerzos le han permitido a la institución lograr importantes reconocimientos como el Premio al Mejor Repositorio Institucional de Colombia⁵ y el Premio Nacional de Ciencia Abierta en su categoría Acceso Abierto, otorgado por Colciencias (ahora Minciencias), en el año 2017.

Este recorrido, acompañado de un trabajo colaborativo y mancomunado de diferentes áreas de la universidad, nos ha permitido avanzar hacia la construcción de una estrategia de materialización de la Ciencia Abierta en la universidad. En la presente ponencia, se describe la forma en que ha sido diseñada esta estrategia y el trabajo que se ha desarrollado hasta el momento para avanzar en su implementación, a través del diseño de una propuesta de Lineamientos Institucionales, la construcción del Portal institucional de Ciencia Abierta, y el desarrollo de una prueba piloto en la Escuela de Medicina y Ciencias de la Salud (EMCS).

Diseño de la propuesta

Siguiendo con el camino trazado, desde 2018, se inicia la reflexión institucional, liderada por el Centro de Recursos para el Aprendizaje y la

⁵ El [Premio al Mejor Repositorio Institucional RENATA 2011](#) reconoce el esfuerzo de las instituciones vinculadas a RENATA por fortalecer la visibilidad de la producción científica, educativa y cultural del país, a través de sus repositorios digitales de acceso abierto.

Investigación (CRAI), sobre la consolidación de una estrategia institucional de Ciencia Abierta. Para ello, se revisaron los modelos existentes y se identificaron aspectos indispensables a incluir en el modelo institucional, gracias a los elementos comunes identificados, que se resumen en el ANEXO 1.

En el año 2019, se conformó un equipo interdisciplinar constituido por miembros de la Vicerrectoría Académica, el CRAI, la Dirección de Investigación e Innovación, la editorial y otras unidades académicas de la universidad, quienes cuentan con conocimientos alrededor de los conceptos que aborda la Ciencia Abierta, con el fin de integrar diversas experiencias y perspectivas.

Este equipo analizó las principales tendencias, buenas prácticas y referentes en el tema a nivel internacional, como el *Marco de Ciencia Abierta propuesto por la Comisión Europea (2018)*⁶ y *La hoja de ruta para un cambio cultural hacia la Ciencia Abierta (2018)*⁷, de la Liga de Universidades Europeas (LERU), entre muchos otros documentos, que permitieron la generación de una base de conocimiento, y una primera propuesta de los pilares de Ciencia Abierta para la universidad.

Como parte de esta primera reflexión, nace un modelo constituido por cinco pilares principales y tres transversales que se muestran en la figura a continuación:



FIGURA 1. PILARES DEL MARCO INSTITUCIONAL DE CIENCIA ABIERTA EN LA UNIVERSIDAD DEL ROSARIO

Fuente: Universidad del Rosario (2020)

Marco de referencia-lineamientos institucionales sobre ciencia abierta

⁶ European Commission. (2018). *Open Science Policy Platform Recommendations*. Publications Office of the European Union.

⁷ League of European Research Universities (LERU). (2018). *Open Science and its Role in Universities: A Roadmap for Cultural Change*.

Para cada uno de estos pilares se propuso una definición y se identificaron los avances que ha logrado la Institución al respecto y los retos a futuro, con lo cual se consolidó una propuesta inicial, que se constituyó como la base para la discusión institucional del tema.

Con el fin de discutir, y enriquecer esta propuesta preliminar, el equipo lideró una reflexión institucional alrededor de Ciencia Abierta, a través de cinco mesas de discusión, una para cada uno de los pilares principales. Las mesas fueron conformadas por diferentes actores de la comunidad universitaria, y tuvieron como objetivo presentar los pilares correspondientes, conocer y discutir alrededor de las prioridades, los intereses y las expectativas de la comunidad alrededor de cada uno de ellos. En total se contó con la participación de 73 miembros de la comunidad, dentro de los cuales se incluyen investigadores, egresados, funcionarios y directivos, quienes de forma democrática y colaborativa compartieron su conocimiento y perspectivas alrededor del tema.

Como resultado de estos esfuerzos, se validó la propuesta de un marco institucional, a través del cual la Universidad del Rosario entrega a la comunidad académica el «aterrizaje práctico» de los lineamientos de ciencia abierta en el contexto de una Institución de Educación Superior (IES) y los retos para su implementación en la universidad. Esta propuesta, se encuentra consolidada en el documento *Marco de referencia-lineamientos institucionales sobre ciencia abierta* para la Universidad del Rosario, en el cual se incluyó, para cada uno de los pilares una definición, los estándares internacionales o premisas que se tendrán en cuenta para su desarrollo, la normativa institucional existente relacionada, los recursos y servicios disponibles para su materialización, y los retos y recomendaciones que van a orientar el desarrollo de las acciones futuras. A continuación, se presenta la definición de cada uno de los pilares (UNIVERSIDAD DEL ROSARIO, 2020):

Comunicación académica en abierto (definición basada en Comisión Europea, 2019, *Future of scholarly publishing and scholarly communication*): para la UR, la comunicación académica en abierto son todas las gestiones encaminadas a generar accesibilidad y usabilidad de los contenidos producidos

por la institución, que garanticen que la investigación llegue a una audiencia más amplia, a partir de una infraestructura que posibilite el fácil acceso y una participación sin restricción. Ello con base en estándares que garanticen accesibilidad y poca o ninguna restricción, la adecuada comunicación, la socialización y la reutilización de los contenidos, con equidad, diversidad e inclusión, así como la responsabilidad de construir sociedad desde la academia.

Datos de investigación en abierto (definición basada en WILKINSON, M. D. et al., 2016; THE OPEN KNOWLEDGE FOUNDATION, 2014 y THE ROYAL SOCIETY, 2012): son datos que se producen en el contexto de la investigación y son dejados a disposición de la comunidad, de forma libre, a través de licencias de uso y cumpliendo con las leyes aplicables a protección de datos personales, con el fin de que terceros puedan visualizarlos, analizarlos y reutilizarlos.

Métricas de nueva generación: son un conjunto de mediciones del impacto de la investigación (aún en construcción), que buscan dar un marco a través del cual 1) se usen de forma responsable las métricas de investigación existentes y 2) se responda a las necesidades de medición y evaluación de la investigación que no están cubiertas por las métricas tradicionales ni las métricas alternativas. Así, las métricas de nueva generación buscan evaluar el impacto de la investigación, no solo a través del número de citas e impacto de las revistas, sino alimentándose de múltiples fuentes que describan diferentes aspectos de la investigación. Son, entonces, métricas multidimensionales (han de incluir mediciones cuantitativas, cualitativas y el concepto de pares expertos), abiertas, transparentes, que no obedecen intereses privados y que buscan medir el impacto más allá de la comunidad académica, hasta los beneficiarios finales de la investigación misma.

Integridad científica: es el marco de principios que orienta a los diferentes actores de la investigación en el cumplimiento de buenas prácticas científicas, desde la creación de un proyecto de investigación hasta el uso de los datos publicados. Incluye: elaboración de acuerdos previos, uso y validación de datos,

derechos y deberes de los participantes, criterios de autoría y divulgación de información confiable.

Ciencia ciudadana: es el conjunto de prácticas de investigación científica en el que confluyen saberes, pensamientos y acciones de distintos agentes sociales, que generan así modelos incluyentes de producción de conocimiento. La ciencia ciudadana: 1) propicia el diálogo de saberes entre la academia y las comunidades; 2) fomenta la investigación inter y transdisciplinar; 3) fortalece el trabajo investigativo mediante conocimientos y procedimientos diversos; 4) involucra múltiples fuentes de información y variadas formas de saber y conocer de los ciudadanos científicos y no científicos, y 5) promueve la ciudadanía activa, la democratización del saber y la búsqueda de soluciones a problemas relevantes para el contexto en el que se desarrollan.

A partir de los pilares definidos, se realizó un análisis de los avances de la institución en cada uno, que permitió construir un panorama actual de los pilares de Ciencia Abierta en la Universidad, el cual se presenta en la figura a continuación:

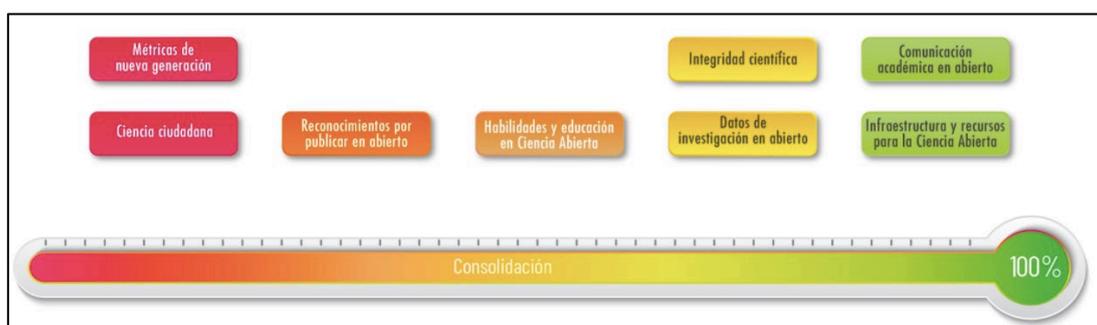


FIGURA 2. PANORAMA DE LOS PILARES DE CIENCIA ABIERTA EN LA UNIVERSIDAD DEL ROSARIO

Universidad del Rosario (2020). Portal institucional de Ciencia Abierta de la Universidad del Rosario: avances UR <https://cienciaabierta.urosario.edu.co/avances-ur.php>

De manera adicional, se desarrolló el Ciclo de Conversaciones sobre Ciencia Abierta, una estrategia de divulgación virtual llevada a cabo través de seis sesiones de Facebook Live, que tuvieron como objetivo abordar el tema de Ciencia Abierta de una forma dinámica y práctica, contribuyendo así a la

sensibilización de la comunidad en torno al tema. Las sesiones fueron lideradas por invitados especiales, expertos, que explicaron cada uno de los pilares de la Ciencia Abierta de manera sencilla y respondieron las inquietudes de la comunidad propiciando la discusión y el intercambio de ideas.

Portal Institucional de Ciencia Abierta

Con el objetivo de visibilizar el compromiso de la UR con Ciencia Abierta, dar a conocer sus avances y la experiencia en la implementación de una estrategia institucional y, al mismo tiempo, contribuir a su promoción en la región, se creó el [Portal Institucional de Ciencia Abierta](#). Siendo un punto de debate e intercambio de experiencias, opiniones y conocimiento alrededor del tema, esta plataforma fue desarrollada de manera conjunta por el equipo interdisciplinar e incluyó las siguientes fases:

- Estructuración del portal web
- Elaboración de contenidos (textos)
- Diseño e implementación
- Divulgación y posicionamiento

El portal da a conocer a la comunidad los avances institucionales alrededor de Ciencia Abierta, sistematiza la bibliografía nacional e internacional relacionada con el tema y se convierte en el punto de encuentro para el intercambio de información, ideas, perspectivas y aportes. Resalta en este sentido la sección Voces de la Comunidad, diseñada para recibir y difundir diferentes experiencias de la comunidad interna y externa, que contribuyen a la discusión y avance del tema a nivel nacional y regional.

Como parte de la estrategia de divulgación del marco institucional propuesto y de posicionamiento del Portal Institucional, se realizó el evento [Perspectivas de la Ciencia Abierta en las Universidades: Aproximación a un modelo institucional](#), con invitados nacionales e internacionales. En este evento, que

contó con la participación de 236 personas de Colombia y otros países de la región se dió a conocer a la comunidad el Marco de referencia-lineamientos institucionales sobre ciencia abierta de la UR y se lanzó el Portal Institucional de Ciencia Abierta.

Prueba piloto

Ciencia Abierta es un tema vivo y relativamente reciente en Colombia, por lo cual se requiere de escenarios institucionales en los cuales se asuma el desafío de llevar la teoría a la práctica, y hacerla realidad en una institución de educación superior.

Con el objetivo de validar los lineamientos institucionales, se ha propuesto desarrollar una prueba piloto en una de las unidades académicas con mayor experiencia en el ámbito de la investigación y generación de conocimiento de la Universidad del Rosario, que es la Escuela de Medicina y Ciencias de la Salud (EMCS).

La fase inicial de este proyecto tuvo como objetivo desarrollar un diagnóstico reflexivo e integrador de los avances y las capacidades actuales de la EMCS en Ciencia Abierta, con el objetivo de construir una base de conocimiento para reconocer el camino recorrido, y plantear las acciones futuras.

El diagnóstico permitió evidenciar los significativos avances de la Escuela, aunque los mayores progresos se han concentrado en algunos de los pilares, como es el caso de la Comunicación académica en abierto e Integridad científica. El primero evidencia un avance contundente, especialmente en el componente de Acceso Abierto, lo cual se ha logrado como resultado de los esfuerzos que ha hecho la institución con la creación de un marco normativo y el desarrollo de diversos recursos y servicios que abarcan la vía verde y la vía dorada. El segundo por la importancia que tiene para la Escuela la definición de lineamientos y buenas prácticas que guíen a sus investigadores en aspectos éticos de sus procesos científicos (UNIVERSIDAD DEL ROSARIO, 2021).

De igual forma, dentro del diagnóstico se identificaron algunos aspectos que aún requieren mayores esfuerzos para lograr su consolidación, dada su novedad en los ámbitos institucional, nacional, e incluso internacional, como es el caso de las métricas de nueva generación, en donde apenas se está trabajando en su definición; la apertura y gestión de los datos de investigación que está iniciando su avance gracias a la estrategia institucional, y la ciencia ciudadana, pilar de gran importancia para la EMCS, dada la relevancia que tiene la inclusión de actores de la sociedad en sus actividades de investigación (UNIVERSIDAD DEL ROSARIO, 2021).

Partiendo de esta base, se diseñó un plan de trabajo para la implementación de una estrategia de apropiación y materialización del modelo de ciencia abierta en la EMCS, con acciones y metas concretas para cada uno de los pilares, que se desarrollarán entre 2021 y 2025. La Escuela será la primera unidad académica en desarrollar esta implementación y servirá de ejemplo a otras unidades académicas de la universidad, por lo cual se espera también que este ejercicio contribuya a una comprensión profunda de los lineamientos institucionales y a su apropiación por parte de la Comunidad Rosarista, así como también aporte a la discusión regional e internacional entorno al tema.

Trabajo futuro en la universidad

El establecimiento de un marco institucional representa, para la Universidad del Rosario, la consolidación de un camino que se ha venido construyendo desde hace más de una década, y a su vez constituye el inicio de un proceso de largo aliento que tendrá como objetivo la materialización de cada uno de los pilares de la Ciencia Abierta en la institución, a través de diversas estrategias y planes de acción. Los mayores retos están relacionados principalmente con dos aspectos; en primer lugar, las nuevas formas y métricas de evaluación, para lo cual la UR realizará un trabajo constante para llegar a un entendimiento correcto de la definición, la medición y el uso de métricas de nueva generación que impulse el fortalecimiento del sistema de evaluación que permita hacer una medición objetiva, transparente, pertinente y replicable del impacto de la

investigación. El segundo gran reto está en la materialización de la ciencia ciudadana, para lo cual se trabajará hacia el reconocimiento de experiencias existentes en la universidad y la generación de un marco institucional que propicie la implementación de la ciencia ciudadana en la UR y el fortalecimiento de los procesos de construcción de confianza con la comunidad.

Como parte de la estrategia orientada a la materialización de la Ciencia Abierta, la universidad continuará con un proceso continuo de promoción y sensibilización de la comunidad, que permita lograr una apropiación profunda del tema, y un reconocimiento de la importancia y los grandes beneficios que tiene la apertura de los resultados de investigación dentro de los objetivos de la institución.

La estrategia de apropiación del modelo en la Escuela de Medicina y Ciencias de la Salud ha permitido evidenciar los avances, las capacidades y las trayectorias que en materia de Ciencia Abierta presenta la EMCS y la universidad en general, a partir de las cuales se han identificado las acciones concretas que se deben contemplar en la unidad como trabajo futuro, enmarcadas en un plan de trabajo para la materialización de los pilares que conforman el modelo institucional. Dentro de este trabajo futuro, se incluye la generación de una cultura hacia la apertura de la investigación científica, la consolidación de alianzas estratégicas y modelos colaborativos, y la evaluación del impacto de la ciencia abierta. La implementación de esta estrategia en la EMCS permitirá llevar el modelo institucional a la práctica, materializar los pilares de Ciencia Abierta en una unidad académica y construir aprendizajes e identificar retos a futuro para la Institución.

Conclusiones y aprendizajes

Ciencia Abierta es un tema amplio, que integra diversidad de elementos y disciplinas, y que además evoluciona continuamente dada su relevancia y novedad a nivel regional e internacional, por lo que construir una estrategia para su promoción y sobre todo para su materialización puede resultar retador y complejo. Por tanto, considerando que el tema involucra diversos frentes, es

posible encontrar un mayor avance en algunos de ellos, como sucede actualmente en la Universidad del Rosario y en el escenario regional, en donde se ha logrado una mayor comprensión y avance en pilares como el acceso abierto y la integridad científica. Así mismo, se identifican algunos aspectos que aún requieren mayores esfuerzos para lograr su fortalecimiento y su consolidación, dada su novedad en los ámbitos institucional, nacional, e incluso internacional, como es el caso de métricas de última generación, datos de investigación en abierto y ciencia ciudadana.

La apuesta institucional por Ciencia Abierta demanda un compromiso decidido y firme por parte de las autoridades, así como esfuerzos constantes, mancomunados y de largo aliento, debido a los cambios que implica en la cultura y procesos institucionales. La Universidad del Rosario ha asumido gustosa este reto y este compromiso, gracias a lo cual ha logrado avances importantes, los cuales han permitido a los miembros de la comunidad tener un conocimiento del tema, percibir y reconocer su importancia y beneficios, así como también generar un gran interés por aprender mucho más sobre el mismo, especialmente desde una perspectiva práctica.

Para el desarrollo de una estrategia disruptiva de este tipo, que cambia las dinámicas tradicionales, es importante identificar las buenas prácticas y experiencias internacionales para aprender de procesos similares y adaptarlos a las características particulares, propias de la Institución y su contexto.

Se recomienda que el proceso de construcción de un marco institucional alrededor de Ciencia Abierta contemple la construcción colectiva involucrando diferentes actores de la comunidad universitaria, con el fin de reconocer sus perspectivas, expectativas y aportes. Esto permitirá considerar los avances y conocimientos existentes en el tema, y trazar una hoja de ruta acorde con las dinámicas propias de la región, la institución y su comunidad.

En conclusión, la Ciencia Abierta es un tema prioritario y su materialización es un desafío para las universidades. Es recomendable aprovechar el trabajo colaborativo entre las instituciones para sensibilizar y consolidar el avance regional, enriquecido con diversas experiencias y perspectivas.

Anexo 1

N°	Fuente	Elementos comunes
1	<u>OECD (2012). Recent policy measures to promote open science. STI policy profiles: strengthening interactions for innovation</u>	<ul style="list-style-type: none"> - Acceso abierto - <i>Datos abiertos</i> - Trabajo colaborativo
2	<u>OECD (2015). Making Open Science a Reality. OECD Science, Technology and Industry Policy Papers, 25</u>	<ul style="list-style-type: none"> - Acceso abierto - <i>Datos abiertos</i> - Revisión por pares abierta - Altmetrics, métrica alternativa de impacto científico - Datos del gobierno abiertos - Ciencia ciudadana
3	<u>COLCIENCIAS. Documento de Trabajo: Ciencia Abierta – Elementos conceptuales</u>	<ul style="list-style-type: none"> - Acceso abierto - <i>Datos abiertos</i> - Investigación abierta
4	<u>European Commission Directorate-General for Research and Innovation. Open Science Monitor</u>	<ul style="list-style-type: none"> - Acceso abierto - <i>Datos abiertos</i> - Comunicación académica abierta
5	<u>REBIUN (2016). Ciencia Abierta: la investigación y los datos científicos accesibles y abiertos a todos los ciudadanos</u>	<ul style="list-style-type: none"> - Acceso abierto - <i>Datos abiertos</i> - Evaluación abierta de resultados de investigación
6	<u>European Commission Directorate-General for Research and Innovation. (2016). Open Innovation, open science, open to the World - a vision for Europe</u>	<ul style="list-style-type: none"> - Acceso abierto - <i>Datos abiertos</i> - Ciencia ciudadana - Código abierto - Pre-impresión - Sistemas alternativos de reputación - Bibliografías colaborativas - Blog de ciencia - Anotación abierta

N°	Fuente	Elementos comunes
		<ul style="list-style-type: none"> - Flujo de trabajo abierto - Data-intensive
7	<u>KRAKER, P. (2011). The case for an open science in technology enhanced learning. <i>Int. J. Technology Enhanced Learning</i>, 3(6)</u>	<ul style="list-style-type: none"> - Acceso abierto - Datos abiertos - Código abierto - Metodología abierta
8	<u>PONTIKA, N., KNOTH, P., CANCELLIERI, M. & PEARCE, S. (2015). Fostering Open Science to Research using a Taxonomy and an eLearning Portal. In: <i>iKnow: 15th International Conference on Knowledge Technologies and Data Driven Business</i>, 21-22 October 2015, Graz, Austria</u>	<ul style="list-style-type: none"> - Acceso abierto - Datos abiertos - Experimentos de investigación abiertos - Métricas e impacto abierto - Evaluación abierta de resultados de investigación - Políticas de ciencia, abiertas - Herramienta de ciencia abierta - Proyectos de ciencia, abiertos

TABLA 1. VISIONES COMUNES DE LA CIENCIA ABIERTA

Elaboración propia

Bibliografía

- EUROPEAN COMMISSION DIRECTORATE-GENERAL FOR RESEARCH AND INNOVATION. (2016). Open Innovation, open science, open to the World - a vision for Europe. <https://ec.europa.eu/digital-single-market/en/news/open-innovation-open-science-open-world-vision-europe>
- EUROPEAN COMMISSION. (2018). Open Science Policy Platform Recommendations. Publications Office of the European Union. <https://op.europa.eu/en/publication-detail/-/publication/5b05b687-907e-11e8-8bc1-01aa75ed71a1>
- KRAKER, P. (2011), The case for an open science in technology enhanced learning. *Int. J. Technology Enhanced Learning*, 3(6).
- LEAGUE OF EUROPEAN RESEARCH UNIVERSITIES (LERU). (2018). Open Science and its Role in Universities: A Roadmap for Cultural Change. <https://www.leru.org/publications/open-science-and-its-role-in-universities-a-roadmap-for-cultural-change>
- OECD. (2012), Recent policy measures to promote open science. STI policy profiles: strengthening interactions for innovation.
- OECD. (2015). Making Open Science a Reality. *OECD Science, Technology and Industry Policy Papers*, 25. https://www.oecd-ilibrary.org/science-and-technology/making-open-science-a-reality_5jrs2f963zs1-en
- PONTIKA, N. et al. (2015). Fostering Open Science to Research using a Taxonomy and an eLearning Portal. In: *iKnow: 15th International Conference on Knowledge Technologies and Data Driven Business*, 21-22 October 2015, Graz, Austria.
- REBIUN. (2016). Ciencia Abierta: la investigación y los datos científicos accesibles y abiertos a todos los ciudadanos. <http://rebiun.xercode.es/xmlui/handle/20.500.11967/70>
- UNESCO. (2020) ¿Qué es Ciencia Abierta? UNESCO lanza consulta global. <https://es.unesco.org/news/que-es-ciencia-abierta-unesco-lanza-consulta-global>
- UNIDAD DE DISEÑO Y EVALUACIÓN DE POLÍTICAS DEPARTAMENTO ADMINISTRATIVO DE CIENCIA, TECNOLOGÍA E INNOVACIÓN – COLCIENCIAS. Documento de Trabajo: Ciencia Abierta – Elementos conceptuales.
- UNIVERSIDAD DEL ROSARIO. (2020) Marco de referencia-lineamientos institucionales sobre ciencia abierta.
- UNIVERSIDAD DEL ROSARIO. (2020). Portal institucional de Ciencia Abierta de la Universidad del Rosario: avances UR. <https://cienciaabierta.urosario.edu.co/avances-ur.php>
- UNIVERSIDAD DEL ROSARIO. (2021). Ciencia abierta: puntos de partida para su comprensión y expansión en la Escuela de Medicina y Ciencias de la Salud.

Capacitação dos Profissionais de informação visando a gestão no Arca – Repositório Institucional da Fundação Oswaldo Cruz

CLAUDETE FERNANDES DE QUEIROZ

Fundação Oswaldo Cruz

claudete.queiroz@iciict.fiocruz.br

ÉDER DE ALMEIDA FREYRE

Fundação Oswaldo Cruz

eder.freyre@iciict.fiocruz.br

RAPHAEL BELCHIOR RODRIGUES

Fundação Oswaldo Cruz

raphael.rodrigues@iciict.fiocruz.br

CATARINA BARRETO MALHEIRO PEREIRA

Fundação Oswaldo Cruz

catarina.barreto@iciict.fiocruz.br

TIAGO MARTINS DA COSTA FERREIRA

Fundação Oswaldo Cruz

Tiago.ferreira@iciict.fiocruz.br

RESUMO

O texto aborda a importância da realização dos trabalhos de sensibilização e capacitação em palestras e cursos ministrados para os profissionais de informação/Bibliotecários que trabalham na gestão do Arca - Repositório Institucional da Fundação Oswaldo Cruz (Brasil), visando assim aprimorar

as competências internas para a melhoria da eficiência, eficácia e qualidade dos documentos inseridos no RI como também disponibilizar o conteúdo em saúde pública para toda a Sociedade.

PALAVRAS-CHAVE

Capacitação profissional; Arca – Repositório Institucional da Fiocruz (Brasil); repositórios; Acesso Aberto; Ciência Aberta.

Introdução

O Repositório Institucional Arca é o principal instrumento de realização do acesso aberto da [Fundação Oswaldo Cruz](#), conforme instituído pela [Política de Acesso Aberto ao Conhecimento](#), tendo como missão «reunir, hospedar, preservar, disponibilizar e dar visibilidade à produção intelectual da Instituição», valorizando assim, o acesso aberto aos documentos (FUNDAÇÃO OSWALDO CRUZ, 2021).

O Arca foi desenvolvido em DSpace, software livre, utilizado por instituições de ensino e pesquisa em todo o mundo, que permite o gerenciamento da produção científica. A criação de uma base de dados, reunindo toda a produção intelectual da Fiocruz, possibilita a preservação e a disponibilização dos documentos dando maior visibilidade e acesso a Sociedade. O Repositório está organizado internamente em Comunidades que representam as unidades técnico científicas da Fiocruz, em Subcomunidades, representando os programas de pós-graduação, e em coleções, que reúnem documentos por tipologia - teses, dissertações, artigos de periódicos, etc. (FUNDAÇÃO OSWALDO CRUZ, 2021).

A Política de Acesso Aberto ao Conhecimento, assim como o Arca contam com uma estrutura de governança, composta por um Comitê da Regulação, um Comitê Gestor e pelos Núcleos de Acesso Aberto ao Conhecimento (NAAC), que trabalham de forma articulada, integrada e complementar. Os NAACs são

formados por profissionais de informação/Bibliotecários que estão envolvidos no âmbito de cada unidade da Fiocruz, sendo responsáveis pela coordenação, gestão, operação, acompanhamento do Repositório. Cabe ao Grupo Gestor do Arca a realização de ações de sensibilização como cursos, treinamentos e apresentação de palestras, realizadas periodicamente tendo como objetivo atender e promover a capacitação dos gestores das Comunidades no Repositório.

O planejamento dessas ações busca manter uma concepção da capacitação em gestão de pessoas, alinhadas com as premissas e referenciais conceituais e metodológicos propostos para o gerenciamento das informações inseridas no Repositório. É importante afirmar que o capital intelectual da Instituição está alinhado às diferentes realidades de trabalho dos profissionais de informação. Desta forma, é fundamental que a gestão das pessoas possa promover competência, organização e responsabilidade com os dados que serão inseridos, de forma a permitir a compreensão do ciclo informacional pertinente aos produtos e serviços disponíveis.

SEGUNDO QUEIROZ E ARAUJO (2020) «o profissional Bibliotecário que atua com Repositórios precisa estar alinhado com outras particularidades, como por exemplo, sua atuação como gestor do sistema, nas comunidades e coleções dentro de um contexto informacional bem abrangente, visando atender a Instituição e seus pesquisadores». Neste sentido, entendemos que as capacitações são fundamentais para o desenvolvimento de um trabalho eficaz para alinhar os conhecimentos técnicos e a troca de experiências.

As competências envolvem também atitudes que são capazes de mobilizar conhecimentos, ações e valores que precisam interagir com as tomadas de decisões e o gerenciamento dos recursos humanos. Os profissionais de informação/Bibliotecários que atuam como gestores, precisam garantir a qualidade e o acesso aos dados que foram depositados (AMANTE, 2014).

Podemos afirmar que o papel do profissional de informação/Bibliotecário para a realização dos trabalhos no Arca é fundamental para o sucesso do Sistema, reconhecido na Instituição como uma ferramenta fundamental e que

oferece a sociedade informações na área de Saúde Pública, alinhadas com o Sistema Único de Saúde (SUS) do Brasil. Os dados produzidos pelos pesquisadores sobre saúde estão disponíveis no Repositório, permitindo assim a disseminação e a preservação da memória institucional para a comunidade científica.

Vale ressaltar que os Repositórios interagem diretamente com a questão do Acesso Aberto e da Ciência Aberta, conteúdos indispensáveis nos treinamentos tendo em vista as demandas dos pesquisadores por conteúdos científicos e dados de pesquisa.

SEGUNDO RODRIGUES (2018), «Ciência Aberta é a atividade científica praticada de modo aberto, colaborativo e transparente, em todos os domínios do conhecimento, desde as ciências fundamentais até às ciências sociais e humanidades. Está associada as temáticas como o acesso aberto, dados abertos, ciência cidadã e sistemas abertos de avaliação pelos pares».

O Acesso Aberto (AA) ou Acesso Livre diz respeito a disponibilização online e sem limitações dos trabalhos resultantes de investigação científica, podendo ser aplicado a todos os tipos de publicações, incluindo artigos científicos, documentos de conferência, teses, etc. (FUNDAÇÃO OSWALDO CRUZ, 2021).

Metodologia

A equipe gestora do Arca, composta por profissionais Bibliotecários, elaborou um Plano de Capacitação para promover cursos, palestras e treinamentos, abordando conceitos sobre Ciência Aberta, Comunicação Científica, Divulgação Científica, Fontes de Informação, Acesso Aberto, Repositórios, Curadoria Digital, Dados de Pesquisa, Preservação Digital, Padrão de Metadados, Direitos Autorais, Autoarquivamento e outros assuntos pertinentes, de forma a oferecer um conteúdo explicativo e de qualidade, com aplicação de exercícios práticos com a interação professor-aluno, emissão de certificados, visando, assim, contribuir para o aprimoramento do fluxo operacional do Repositório.

A metodologia proposta envolveu primeiramente a formatação de treinamentos presenciais sobre o Repositório Arca e cursos ligados aos assuntos correlatos e do estabelecimento de critérios, padrões e procedimentos que atendessem a demanda dos gestores de informação das unidades da Fiocruz, visando fornecer a orientação para a correta inserção dos registros e realização de curadoria. Os treinamentos e cursos foram divididos em módulos com a descrição do conteúdo programático, dentro de uma carga horária definida adotando uma estratégia direcionada ao público alvo, desde a contextualização histórica dos repositórios até a explanação sobre o Arca. Cabe ressaltar que ao final dos cursos, é oferecido um formulário de avaliação onde o aluno pode informar os pontos positivos, negativos e ainda dar sugestões ou fazer comentários. Este documento é tabulado para a descrição dos itens que precisarão ser melhorados, atualizados ou até descartados. Apresentamos nas FIGURAS 1, 2 e 3 os resultados gerados das tabulações dos dados referentes aos pontos positivos, negativos e das sugestões e comentários.



FIGURA 1: TABULAÇÃO DE DADOS – TREINAMENTOS: PONTOS POSITIVOS



FIGURA 2: TABULAÇÃO DE DADOS – TREINAMENTOS: PONTOS NEGATIVOS

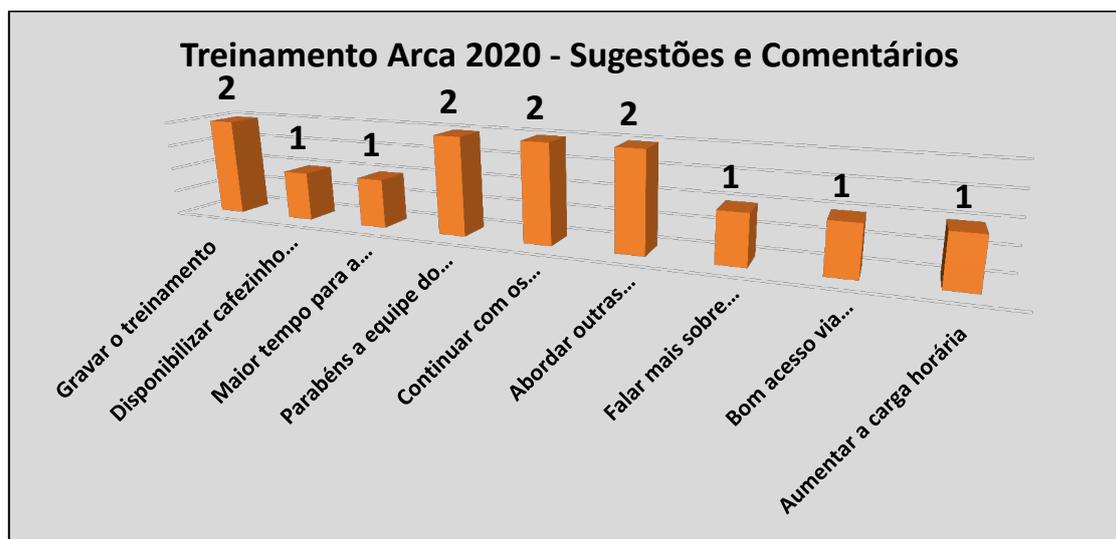


FIGURA 3: TABULAÇÃO DE DADOS – TREINAMENTOS: SUGESTÕES E COMENTÁRIOS

Após a crise da pandemia da COVID-19, foi necessário redesenhar esse processo, tendo em vista uma nova realidade mundial, que passou a contar com a ministração das aulas via webconferência, tendo a mesma formatação dos cursos presenciais, mas com a interação online dos alunos.

O *Manual de preenchimento de metadados no Arca* foi desenvolvido como uma instrumento para ajudar e orientar os alunos no preenchimento da entrada dos documentos, estabelecendo consistência no tratamento dos metadados. O

manual foi organizado seguindo a ordem alfabética das tipologías/coleções disponíveis no Repositório, contendo a descrição e exemplos em cada uma delas, além de novos tutoriais. O manual está disponível em acesso aberto no [link](#).

Justificativa

Tendo em vista que a alimentação no Repositório Arca é descentralizada, sendo realizada através das diversas Unidades da Fiocruz e de suas Bibliotecas, tornou-se fundamental o monitoramento e a gestão da qualidade dos dados preenchidos e inseridos. Dentro deste contexto as aulas se tornaram uma preciosa ferramenta para firmar padrões e procedimentos de alimentação das informações e dos objetos digitais. As aulas complementam uma lacuna que existe no que se refere aos assuntos ligados diretamente ao Repositório, alinhando assim, diretrizes e procedimentos com um padrão de eficácia, além de propiciar uma segurança para os gestores do Arca na inserção dos documentos (MARANHÃO, QUEIROZ & BELCHIOR, 2017).

Desta forma, a capacitação do profissional de informação/Bibliotecário, enquanto gestor do conhecimento, deve compreender o potencial dos sistemas de informações digitais em rede dentro das bibliotecas e deve colaborar na promoção e uso do acesso aberto, promovendo e disponibilizando os documentos publicados, bem como identificar os recursos de informação externos indispensáveis para que a organização e seus membros possam desenvolver as suas atividades (AMANTE, 2014).

Resultados

A sistematização dos treinamentos, curso e palestras, destacam a importância do papel do Bibliotecário na gestão das informações relevantes para o campo da Saúde e Pesquisa dentro da Fiocruz. Podemos afirmar, portanto, que este trabalho pode propiciar, de forma colaborativa, a melhoria na qualidade dos metadados e objetos digitais armazenados, além de promover a

visualização das informações e garantir a recuperação mais precisa para os usuarios.

No período de 2016 a 2020 foram treinadas 225 pessoas com a realização de 38 treinamentos, ressaltando a importancia das capacitações internas conforme visto na FIGURA 4.

As tarefas requeridas e apresentadas até aqui evidenciam o importante papel dos bibliotecários nos Repositórios, que estão capacitados para lidar com o cotidiano da gestão das informações bibliográficas, dos dados de pesquisa, da curadoria e também da Preservação Digital. O imenso desafio e as responsabilidades deste profissional, se torna evidente e ascende para a qualidade na formação diversificada nas diferentes áreas de atuação do conhecimento dos gestores do Repositório. Esse profissional com perfil de gestor, ao se capacitar, passa a priorizar processos, critérios e estratégias para sistematizar e organizar as informações, valorizando o conhecimento adquirido no trabalho e nas trocas de experiências durante as aulas.

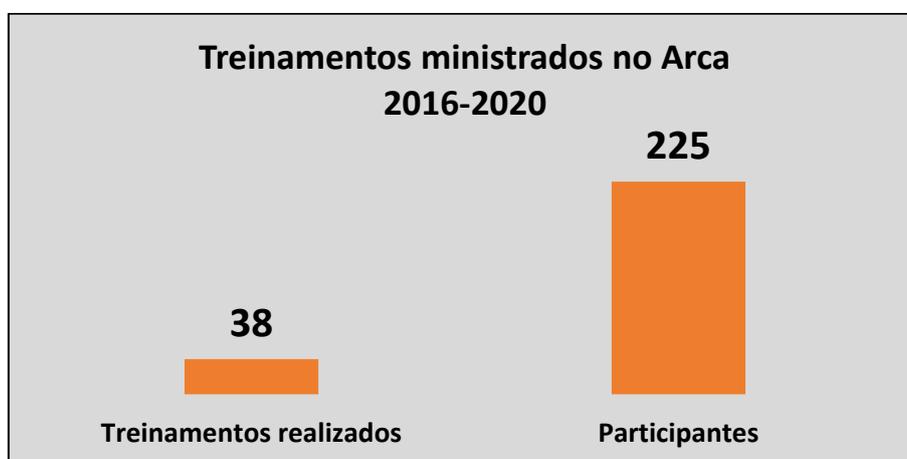


FIGURA 4: TREINAMENTOS MINISTRADOS PELA EQUIPE GESTORA DO ARCA

Conclusão

Concluimos este trabalho reafirmando a importância da capacitação informacional do profissional de informação/Bibliotecário que trabalha em Repositórios, considerando a sua atuação fundamental para o gerenciamento e

sucesso desses sistemas. O importante é que eles podem interagir principalmente com o usuário e com o conhecimento produzido, de forma a garantir a confiabilidade, o acesso e a integridade das informações. Esse gestor ao se capacitar passa a priorizar processos, critérios e estratégias para sistematizar e organizar as informações, valorizando assim o conhecimento adquirido no trabalho durante as aulas.

Bibliografía

- AMANTE, M. J. (2014). O bibliotecário como gestor do conhecimento: o caso dos repositórios. *RECIIS - Revista Eletrônica de Comunicação, Informação e Inovação em Saúde*, 8(2), 243-254. <https://www.arca.fiocruz.br/handle/icict/17100>
- FARIAS SALES, L. et al. (2019). Competências dos bibliotecários na gestão dos dados de pesquisa. *Ciência da Informação*, 48(3), 303-313. <https://www.arca.fiocruz.br/handle/icict/43074>
- FUNDAÇÃO OSWALDO CRUZ. (2021). *Sobre o Arca*. <https://www.arca.fiocruz.br/terms/sobre.jsp>
- MARANHÃO, A. M. N.; NEVES DE QUEIROZ, C. F.; BELCHIOR RODRIGUES, R. (2017). Curadoria digital de dados no Arca - Repositório Institucional da Fiocruz: relato de experiência. *RECIIS - Revista Eletrônica de Comunicação, Informação e Inovação em Saúde*, 11, 1-4. <https://www.arca.fiocruz.br/handle/icict/23725>
- RODRIGUES, E. et al. (2018). *O que é Ciência Aberta?* Rio de Janeiro: Campus Virtual Fiocruz/Escola Corporativa Fiocruz; Universidade do Minho. Material multimídia (Microcurso 1/3, Série 1 - Formação Modular em Ciência Aberta).

Mapeamento dos projetos brasileiros em Ciência Cidadã e Biodiversidade

LAURA REGINA DO CANTO LEAL

Universidade Federal do Rio Grande do Sul

laura0806@bol.com.br

<https://orcid.org/0000-0002-0631-5607>

LEILA MORÁS SILVA

Universidade Federal do Rio Grande do Sul

moras.leila@gmail.com

<https://orcid.org/0000-0003-4694-7702>

RENE FAUSTINO GABRIEL JUNIOR

Universidade Federal do Rio Grande do Sul

rene.gabriel@ufrgs.br

<http://orcid.org/0000-0003-1021-3360>

RESUMO

O presente estudo exploratório tem como objetivo mapear e identificar projetos de Ciência Cidadã aplicados à biodiversidade no Brasil por meio da plataforma do Sistema de Informação sobre a Biodiversidade Brasileira (SiBBr). Destaca-se alguns projetos a partir da sua relevância, dimensão, participação de cidadãos voluntários, origem e atuação. Foram identificados um total de 34 projetos de biodiversidade na SiBBr no Brasil, sendo que 13 desses projetos foram selecionados para este estudo. Conclui-se que é visível a diversidade dos assuntos abordados nos projetos desenvolvidos e todo o engajamento dos pesquisadores e voluntários, mas há muito a ser explorado quanto à transparência das informações coletadas, pois os projetos são incipientes frente a todo um

emaranhado de informações e dados a serem investigados sobre Ciência Cidadã e biodiversidade no Brasil.

PALAVRAS-CHAVE

Biodiversidade; Ciência Cidadã; projetos; voluntários; biodiversity; Citizen Science; projects; volunteers.

Introdução

A ciência aberta¹ está mudando o modelo da prática científica, que em consonância com o desenvolvimento da cultura digital está disponibilizando mais informações tanto para pesquisadores como para a sociedade. Desta forma, o fazer ciência torna-se um desafio face às rápidas transformações tecnológicas que vivemos na atualidade. As diferentes demandas das comunidades e os diferentes espaços de construção do saber, modificam as relações entre a sociedade e os meios científicos.

O movimento de Ciência Aberta tem seu início em 1991, principalmente no combate abusivo das assinaturas de bases de dados, um de seus precursores foi com Paul Ginsparg, quando propõe o desenvolvimento de servidores de e-print, o arXiv, repositório de preprints de acesso aberto, voltado para áreas da física, matemática, entre outras (ALVAREZ & CAREGNATO, 2017).

No entanto, é somente em 2002 que toma corpo com a reunião de pesquisadores em prol da Ciência Aberta, com o manifesto da Iniciativa do **Acesso Aberto de Budapeste** (BOAI). Este documento foi o início para as pesquisas revisadas por pares em revista de acesso aberto, buscando

¹ Ciência aberta - movimento da comunidade acadêmica de tornar a pesquisa científica acessível para todos. É um comprometimento vindo dos pesquisadores para com o público para disseminar livre e abertamente os *inputs* e *outputs* das pesquisas científicas de uma forma compreensível, acessível e aproveitável, e que permita a reprodutibilidade (CASTRO MARTINS, 2020).

dinamizar e disponibilizar de forma gratuita o acesso de todos os usuários ao material produzido.

A Ciência Aberta ao englobar o conceito dos oito pilares apresentados na [Plataforma de Políticas de Ciência Aberta](#) (Open Science Policy Platform) - acesso aberto, dados abertos, revisão por pares, software de código aberto, cadernos abertos de laboratórios, recursos educacionais abertos, redes sociais científicas e acadêmicas e ciência cidadã.

O foco deste estudo está em um destes pilares, o da ciência cidadã, que se dá uma das interações entre o conhecimento acadêmico e o não acadêmico ou popular, com iniciativa de colaboradores de dentro e de fora do meio científico, contribuindo com uma ciência colaborativa entre cientistas e não cientistas, chamada Ciência Cidadã.

Para PARRA (2015), a Ciência Cidadã refere-se a uma «ciência de colaboração», ou seja, sua dinâmica de produção na coleta de dados, informações e métodos aplicados em seu processo é entendida como uma forma de democratizar a ciência, onde cidadãos voluntários se apropriam ou criam conhecimento científico, podendo intervir no processo de tomada de decisão de políticas públicas, envolvendo temas de ciência e tecnologia, e levando à produção de conhecimento em ambientes externos à comunidade científica.

No que se refere a participação e contribuição dos voluntários ou cidadãos não acadêmicos nas pesquisas e projetos, YAMAMOTO (2020) destaca que estes desempenham uma atividade de aprendizado ou lazer por meio da qual colaboram formalmente com um projeto de pesquisa, cujo resultado pode retornar para o participante na forma de melhoria das condições sociais, ambientais ou até na elaboração de novas políticas públicas.

Ressalta-se que a interação nas pesquisas entre voluntários e cientistas profissionais são desenvolvidas com metodologia científica específicas para cada projeto, nada é efetuado de forma aleatória, pois pode gerar discrepância nos dados registrados. A contribuição dos voluntários agrega não somente nos resultados das pesquisas científicas, mas também suas participações podem

contribuir em benefícios na qualidade de vida da sociedade e do próprio meio ambiente.

Embora a Ciência Cidadã ainda seja incipiente no Brasil, ela retrata o interesse mútuo do voluntário ao cientista com o intuito de criar estratégias e oportunidades para o desenvolvimento sustentável do planeta. Diante desse contexto, temos a Ciência Cidadã na área da Biodiversidade que vem sendo estudada, aplicada e publicada, demonstrando sua importância como metodologia de pesquisa científica e contribuição da ciência para sociedade, e ao mesmo tempo, de educação científica (NASCIMENTO, 2018: 27).

Sob essas perspectivas, este trabalho tem por objetivos: analisar projetos desenvolvidos no Brasil de Ciência Cidadã em biodiversidade, tendo como elementos de análise averiguar como ocorre a contribuição e a participação dos voluntários junto com os cientistas profissionais nos projetos elencados neste estudo.

A pesquisa se caracteriza como exploratória, a fonte de informação foi o Sistema de Informação sobre a Biodiversidade Brasileira (SiBBr), que tem indexado 146 instituições, 340 coleções, 512 conjuntos de dados e 16.695.657 ocorrências. Também pesquisou-se no site de cada projeto ou em suas páginas nas redes sociais (Facebook), quando existente, para complemento de informações. Para definir o corpus de análise foi utilizado uma amostragem aleatória simples, do tipo probabilística. Para contextualização do estudo, parte-se de uma revisão de literatura sobre o tema.

Ciência Cidadã e a Biodiversidade

A Ciência Cidadã está integrada em várias áreas do conhecimento como na saúde, antropologia, astronomia, biodiversidade, entre outras, desenvolvendo diferentes tipos de projetos. Observando os projetos existentes sobre Ciência Cidadã, verificou-se que os mais explorados são os vinculados à biodiversidade, pois enfrentam os desafios do desenvolvimento sustentável e igualitário.

De acordo com o SiBBR (online), a Ciência Cidadã é «baseada na parceria entre cientistas e cidadãos que, trabalhando juntos, conseguem otimizar a produção científica». Essa parceria busca coletar dados e informações que serão analisados por um cientista. Tudo isso gera uma melhor compreensão do cidadão que participou como voluntário a respeito da importância do processo científico na produção do conhecimento, e para o cientista profissional, uma melhor qualidade na coleta dos dados.

Ao analisar a situação da biodiversidade em relação às aves no Brasil, MENEGASSI (2021), faz um recorte e afirma que sete aves da Mata Atlântica foram extintas nas últimas décadas e outras nove estão ameaçadas de seguir o mesmo caminho se esforços de conservação não forem realizados. Relata também que o Brasil é o segundo país no mundo em números de aves ameaçadas de extinção, e que a maior parte delas se encontra na Mata Atlântica. Aponta que para a sobrevivência das aves se faz necessário «investimentos contínuos e ampliados em advocacy [políticas públicas de defesa], proteção de habitat, restauração, gestão intensiva de populações, pesquisas direcionadas e engajamento público». No entanto, reporta que nem tudo está perdido e que as taxas de desmatamento da Mata Atlântica vêm caindo com ações de recuperação e restauração florestal, face, em parte, ao engajamento de pessoas comuns, interessadas nas aves, que participam com dedicação, colaboração e recursos em projetos de Ciência Cidadã. Portanto, nesse cenário, a Ciência Cidadã é um recurso, senão uma ferramenta, que vem ao encontro do auxílio no conhecimento da biodiversidade.

Tem-se como um dos mais antigos projetos de Ciência Cidadã, em biodiversidade, o Christmas Bird Count, no final do século XIX. Havia uma preocupação dos cientistas e naturalistas, da época com o declínio da população de pássaros devido à caça lateral² dos mesmos. O ornitologista Frank Chapman, da incipiente Audubon Society (EUA), no natal de 1900, propôs uma nova tradição de férias, que ao invés da caça lateral (Christmas Side Hunt), fosse efetuada a contagem e registro das aves, separadas por espécie, durante

² Caça lateral - escolha do lado onde iriam caçar.

as férias. Na época participaram 27 observadores, em 25 locais diferentes e registraram no total 90 espécies em contagens combinadas. O projeto permanece ativo até o presente momento e é no período de 15 de dezembro a 5 de janeiro de cada ano que a mágica acontece, quando milhares de voluntários participam da coleta de dados, que serão usados para avaliar a saúde e ações que serão tomadas para a conservação das populações de pássaros (Audubon, online).

Conforme TEIXEIRA, COSTA & ZAHER (2015), o marco inicial da Ciência Cidadã no Brasil, foi o «recebimento de serpentes pelo Instituto Butantan em 1911, prática iniciada por Vital Brazil»³, que utilizava-se da permuta de cobras por soro antiofídico. Portanto, tem-se na figura de Vital Brazil como o pioneiro em Ciência Cidadã brasileira, pois contava com a participação da população nas permutas, além de torná-la sua aliada na prática científica, assim, conseguia divulgar os feitos científicos da instituição dando mérito a todos os participantes.

No Brasil, a primeira plataforma criada para integrar os projetos de biodiversidade foi a Sistema de Informação sobre a Biodiversidade Brasileira (SiBBR), desenvolvida pelo Ministério da Ciência, Tecnologia e Inovação (MCTI), com suporte técnico da ONU Meio Ambiente (UNEP) e apoio financeiro do Fundo Global para o Meio Ambiente (GEF), com objetivo de «reunir e dar acesso a dados e informações para subsidiar políticas públicas, apoiar ações de conservação e de uso sustentável bem como promover o conhecimento da nossa biodiversidade» (SiBBR, online).

A Biodiversidade, é um dos elementos centrais para o desenvolvimento da humanidade e responsável pelo equilíbrio ambiental, vem sendo cada vez mais reconhecida e valorizada por sua gama de benefícios na área socioeconômica como em matéria-prima, na medicina ou na indústria. As boas práticas de desenvolvimento sustentável refletem no contexto econômico, político, social e ambiental da sustentabilidade, potencializando ainda mais a importância de ações de cuidados à biodiversidade.

³ Cientista e primeiro diretor do Instituto Butantã.

Para VILLAR (2020), «a biodiversidade, ou a diversidade biológica, é o conjunto de todos os seres vivos existentes, o que inclui todas as plantas, animais e microorganismos da Terra». Com essa definição podemos dimensionar a importância da conscientização da preservação e do investimento em projetos voltados à biodiversidade.

Para a Convenção sobre Diversidade Biológica⁴ (Brasil, 2020), a biodiversidade é entendida como a variabilidade de organismos vivos de todas as origens, compreendendo, dentre outros, os ecossistemas terrestres, marinhos e outros ecossistemas aquáticos e os complexos ecológicos de que fazem parte; compreendendo ainda a diversidade dentro de espécies, entre espécies e de ecossistemas. Baseando-se neste conceito diversificado de espécies que engloba o assunto, é de extrema relevância destacar a inclusão da espécie humana como um dos principais componentes do sistema como dependente dos serviços e bens ambientais oferecidos pela diversidade biológica natural.

Em reportagem para a BBC Brasil, McGRATH (2019), aborda que segundo a ONU cerca de «1 milhão de espécies de animais e vegetais estão ameaçados de extinção» e que a maior parte disso, devido a ação humana. Além disso, no Brasil, a Amazônia, a floresta mais biodiversa do mundo, vem sofrendo impactos severos devido ao avanço brutal do desmatamento. De acordo com Villar, R. (2020), no período «[...] de agosto de 2019 a abril de 2020, os alertas de desmatamento indicam a perda de 5.483 km² de floresta», sendo esse «o maior índice dos últimos cinco anos».

Dentre os objetivos traçados pelas Nações Unidas, no que se refere aos Objetivos de Desenvolvimento Sustentável (ODS), da Agenda 2030⁵ é buscar promover o manejo sustentável das florestas, o combate à desertificação revertendo a degradação da terra e interrompendo o processo de perda da biodiversidade. Essa iniciativa representa ainda mais a importância desse tipo

⁴ Convenção sobre Diversidade Biológica (CDB) é um tratado da Organização das Nações Unidas e um dos mais importantes instrumentos internacionais relacionados ao meio ambiente.

⁵ Plataforma Agenda 2030: <http://www.agenda2030.org.br>

de proposta para o desenvolvimento sustentável dos ecossistemas e para toda a humanidade, buscando implantar projetos voltados à conscientização de pesquisas a partir de uma ciência mais humana, com experimentos propostos pensados na população pesquisada, a Ciência Cidadã.

Os projetos de Ciência Cidadã, possibilitam, aos cidadãos voluntários, conhecimento científico sobre conservação, preservação, cidadania, entre outros, e em contrapartida proporciona aos cientistas informações que não teriam acesso rápido se não existisse o compartilhamento, por exemplo, do monitoramento e coleta dados efetuadas que são efetuadas por esse grande grupo. Os projetos podem ser desenvolvidos por um tempo determinado, com data de início e de fim, ou então por um tempo indeterminado, sem data prevista para ser encerrado/concluído.

Resultados da pesquisa

Foram identificados um total de 34 projetos de biodiversidade na SiBBr no Brasil, com base nesta tabela, foi aplicado a tabela de números aleatórios para realização da leitura e caracterização dos projetos, sendo que 13 desses projetos foram selecionados para este estudo. A escolha da plataforma do SiBBr para a pesquisa deve-se pelo fato de ela oferecer diversos projetos em ciência cidadã voltados para a biodiversidade brasileira em um único local. Além disso, foi necessário buscar algumas informações para complemento das pesquisadas nos sites de cada projeto ou em suas páginas nas redes sociais, Facebook, quando existentes, por disponibilizarem maiores detalhes e informações a respeito.

Quanto aos projetos selecionados para este estudo, destacam-se os de relevância quanto a sua dimensão, participação de cidadãos voluntários e de origem e atuação em território nacional, além da importância e atuação dos que se encontram concluídos. Levou-se também em consideração ao fazer a escolha, a diversidade da área dos projetos, propósitos e visibilidade a nível Brasil. Pode-se verificar que alguns projetos indexados na plataforma do SiBBr não estão atualizados, não localizando um motivo que justifique.

Com base no resultado das buscas no SiBBR, foram analisados os projetos em biodiversidade, assim como a forma como os não cientistas interagem com a ciência cidadã por meio da participação e seu desempenho nos projetos. O QUADRO 1 sistematiza os projetos analisados, seu contexto de aplicação e a relação destacada com a Ciência Cidadã.

Período de atuação	Instituição de atuação	Características gerais dos projetos	Características destacadas em Ciência Cidadã
Atlas de Registros de Aves Brasileiras (ARA)			
Ativo. Não apresenta a data de início do projeto, somente lista conclusão a do primeiro ciclo em 2014, e início do segundo ciclo teve seu início em 2017.	Governamental: CEMAVE (Centro Nacional de Pesquisa e Conservação de Aves Silvestres). Os primeiros estudos da CEMAVE sobre registros de aves, datam da década de 1970.	O ARA, é um dos projetos ativos mais antigos do país. Tem como proposta reunir e organizar informações sobre a distribuição das aves silvestres no território brasileiro, facilitando o acesso aos dados sobre as aves. As informações coletadas e disponibilizadas publicamente, tem por finalidade aumentar o conhecimento sobre as aves e servir de base para preservação das espécies. O projeto apresentou, até o momento, 300 mil registros no registro de 1859 espécies, o que corresponde a aproximadamente 97% da avifauna brasileira.	A participação dos voluntários é na colaboração por meio de observações, monitoramento e do envio de dados de registro das aves.
Wikiaves			

Período de atuação	Instituição de atuação	Características gerais dos projetos	Características destacadas em Ciência Cidadã
Ativo. Início em 2008.	Criador e administrador do projeto: Reginaldo Guedes.	De conteúdo interativo, tem por objetivo apoiar, divulgar e promover a observação de aves em território nacional. Alberga por volta de 300 mil registros ano, sendo a maioria de fotos. Conta com cerca de 27 mil usuários, 2 milhões de fotos, 120 mil sons, 1878 mil espécies e 8 mil visitantes diários. É o projeto que detém o maior número de informações, via internet, no país.	A participação dos voluntários é por meio de ferramentas no controle de fotos, sons, textos, identificação de espécies, formação de mapas, comunicação entre observadores, entre outros.
AE Trapp			
Ativo Desde 2015.	O idealizador e Coordenador Tecnológico e Científico é Oda Scatolini. Muitas das pesquisas são efetuadas em parceria com a Prefeitura Municipal de Recife e com o Inciti/UFPE.	O projeto tem por objetivo engajar a população no monitoramento de populações de mosquitos Aedes, vetores de zika, dengue, chikungunya e febre amarela urbana, assim como desenvolver uma solução na adaptação de ovitrampas ⁶ para ser utilizada por voluntários. Os dados obtidos permitem que as comunidades e agentes públicos possam detectar os	Os voluntários têm como recurso um aplicativo para Android e IOS que possibilita fotografar amostras dos locais de monitoramento e enviá-las a um servidor, onde um software faz a contagem de ovos presentes em cada amostra.

⁶ Ovitrapas são iscas simples, que simulam um ambiente propício para a deposição de ovos de mosquitos do gênero *Aedes*.

Período de atuação	Instituição de atuação	Características gerais dos projetos	Características destacadas em Ciência Cidadã
		focos de vetores, fazer comparativos, analisar e formular estratégias para o combate dos mosquitos.	
Sistema de Informação em Saúde Silvestre (SISS-Geo)			
Ativo. Desde 2007.	<p> FIOCRUZ (Fundação Oswaldo Cruz) integrado ao Projeto de Nacional de Ações Público-privada para a Biodiversidade (PROBIO II); financiado pelo Global Environmental Facility (GEF), Banco Mundial, Fundo Brasileiro da Biodiversidade e Caixa Econômica Federal. </p>	<p> O projeto faz o monitoramento da fauna brasileira em tempo real, focando na saúde dos animais silvestres e no cuidado de doenças cujos agentes infecciosos circulam entre animais e humanos, como as zoonoses. Para a coleta de dados é utilizado aplicativos móveis. Os registros são fundamentais, pois contribuem para a conservação da biodiversidade e a qualidade de vida dos animais, assim como torna possível prever novas doenças de risco para espécies silvestres, domésticas e humanos. </p>	<p> Os voluntários participam monitorando a fauna fotografando e reportando as condições físicas dos animais e do ambiente em seu entorno. </p>
Biofaces			

Período de atuação	Instituição de atuação	Características gerais dos projetos	Características destacadas em Ciência Cidadã
Ativo. Não há informações sobre seu início	Não há informações	O objetivo do projeto é aproximar as pessoas com a natureza, estimulando a Ciência Cidadã. Os colaboradores ganham visibilidade, por meio de um ranking, onde lhes é dado crédito das imagens armazenadas. São cerca de 5296 participantes em todo o Brasil, com registros de 16423 espécies de animais, distribuídos anfíbios (595), pássaros (4088), peixes (752), invertebrados (9403), mamíferos (856) e répteis (696).	Os voluntários compartilham os dados da vida selvagem por meio de fotos, vídeos, desenhos, listas de registro, áudios (vocalização) ou artigos científicos de uma ou mais espécies selvagens.
Guardiões da Chapada			
Ativo. Desde de 2015.	Iniciativa de professores e estudantes da Universidade Federal da Bahia vinculado ao Instituto Nacional de Ciência e Tecnologia - INTREE Tem como principais parceiros os Polinizadores do Brasil, Ministério do Meio Ambiente.	Esse projeto se desenvolveu a partir do conhecimento gerado sobre a interação planta-polinizador, onde os esforços possam ser direcionados para que seja possível contribuir com a conservação da natureza, por meio da polarização. O número de participantes ainda é pequeno, são 11 voluntários, com 55 observações de 13 espécies.	Os voluntários colaboram no envio de fotos da interação flor-visitante floral, de qualquer lugar da região da Chapada da Diamantina. É necessário que o voluntário informe a data, horário e local onde a foto foi retirada.

Período de atuação	Instituição de atuação	Características gerais dos projetos	Características destacadas em Ciência Cidadã
Ecoa			
Encerrado. Iniciou em novembro de 2015 e foi concluído em final de 2016.	Coordenado por Rafael Chiaravalloti, tendo como principal financiador o Fundo Itaú Ecomudança/Ekos Brasil.	O projeto Ecoa, foi desenvolvido no período de 12 meses com o objetivo voltado para o reflorestamento de regiões denominadas APP (Áreas de Preservação Permanente), em quinze assentamentos rurais de Mato Grosso do Sul, tendo como meta a redução na quantidade de carbono na atmosfera e aumentar a eficiência hídrica dos locais atendidos. O projeto foi estendido para o mapeamento de árvores frutíferas importantes para a economia local como Baru (<i>Dipteryx</i> sp.), Jatobá (<i>Hymenaea</i> sp.) e Bocaiúva (<i>Acrocomia</i> Sp.). Nos 12 meses de projeto, foram semeados mais de 20 quilos de sementes de espécies nativas do cerrado e plantadas mais de 500 mudas de outras espécies, assim como a construção de mais de 5 mil metros lineares de cerca e de 5 açudes.	A participação dos voluntários ocorre por meio do aplicativo ExCities, onde reportam o desenvolvimento das plântulas e das sementes, fazem o acompanhamento do reflorestamento e contribuem com informações se as ações desenvolvidas no projetos estão ocorrendo dentro do esperado ou se estão apresentando problemas como por exemplo o aparecimento de animais nas áreas de restauração.
Projeto Exoss			

Período de atuação	Instituição de atuação	Características gerais dos projetos	Características destacadas em Ciência Cidadã
Ativo. Início em maio de 2015.	Deriva da união entre astrônomos profissionais e amadores e estão ligados ao observatório nacional, sob a coordenação de Marcelo De Cicco.	O projeto EXOSS, é voltado para o estudo de meteoros e bólidos, tendo como objetivo o registro de meteoros e o desenvolvimento de estudos astrométricos e astrofísicos; além de catalogar novos riantes do hemisfério sul; colaborar com instituições de ensino e pesquisas contribuindo com a sociedade e incentivar a produção de artigos científicos. Atualmente (05/2021), são 60 estações de monitoramento, em 13 estados, com 60 associados distribuídos entre 19 instituições.	Os voluntários efetuam a captura e análise de dados utilizando metodologia científica, sob orientação de astrônomos profissionais, visando a precisão e exatidão de dados e informações, por meio de videomonitoramento do céu noturno.
Onde estão as Baleias e Golfinhos?			
Ativo. Desde 2013.	Conta com o apoio do Instituto Mar Adentro, uma associação civil sem fins lucrativos.	Os desenvolvedores do projeto visam mobilizar e envolver a sociedade no geral, na pesquisa científica, por meio da colaboração entre voluntários e cientistas. Com os dados registrados é possível elaborar o mapeamento da biodiversidade e distribuição de cetáceos na cidade do	O papel dos voluntários é reportar os avistamentos de baleias e golfinhos, por meio de fotos e/ou vídeos de cetáceos nas águas costeiras da cidade do Rio de Janeiro.

Período de atuação	Instituição de atuação	Características gerais dos projetos	Características destacadas em Ciência Cidadã
		<p>Rio de Janeiro. Contaram com 5.826 voluntários e 292 registros de avistamentos no Rio de Janeiro no período de outubro de 2013 a fevereiro de 2018.</p>	
MIND.Funga			
<p>Ativo. Desde de 2019.</p>	<p>Coordenado pelo Laboratório de Micologia da Universidade Federal de Santa Catarina, MICOLAB/UFSC.</p>	<p>O projeto é coordenado pelo Laboratório de Micologia da Universidade Federal de Santa Catarina (MICOLAB/UFSC). Tem por objetivo realizar o monitoramento de macrofungos de altitude. O projeto visa também capacitar e sensibilizar os voluntários sobre a importância do reino fungi para a conservação da biodiversidade e das espécies de macrofungos, principalmente espécies endêmicas e ameaçadas de extinção. Busca por meio da Ciência Cidadã ampliar a participação e o envolvimento da sociedade em ações voltadas à conservação da biodiversidade das unidades de</p>	<p>Os voluntários participam enviando fotos de espécies de macrofungos, alimentando, assim, um banco de dados a partir desse envio de imagens. Os registros se dão por meio da utilização de aplicativo de celular, desenvolvido para esse fim.</p>

Período de atuação	Instituição de atuação	Características gerais dos projetos	Características destacadas em Ciência Cidadã
conservação.			
City Nature Challenge 2021: Grande Curitiba, Brasil			
Ativo. Desde abril de 2020, de caráter anual.	Responsável: Bio Tiba - Projetos de Biodiversidade. Tem na Universidade Tecnológica Federal do Paraná (UTFPR) a principal instituição articuladora dos eventos.	O projeto foi desenvolvido para a cidade de Curitiba e região metropolitana, de caráter anual, em um curto período de tempo, entre final de abril e início de maio e no final de setembro e início de outubro e tem por objetivo os registros da biodiversidade urbana e a busca de uma interação maior entre seres humanos e a natureza urbana, no outono e primavera. Os registros são identificados por especialistas até o menor nível taxonômico possível. Para os registros de fotos e sons é utilizado aplicativo próprio da iNaturalist. Foram efetuadas 3786 observações, de 974 espécies e contou com 179 voluntários.	A participação dos voluntários é na observação, coleta e registro de imagens e sons das espécies de animais urbanos, exceto gatos e cachorros domésticos, sendo que os participantes recebem os créditos de seus achados nas fotos publicadas na plataforma.
Blue Change: Bebeu água?			
Encerrado. Ano de 2017.	Instituição desenvolvedora do projeto Universidade Federal Fluminense	Projeto desenvolvido tendo como proposta contribuir com a educação e a conscientização dos alunos de ensino	A participação dos alunos voluntários foi na coleta dos dados por meio de entrevistas sobre a forma como as

Período de atuação	Instituição de atuação	Características gerais dos projetos	Características destacadas em Ciência Cidadã
	(LATEC/UFF) e Universidade Federal do Rio de Janeiro (UFRJ).	médio da Escola Santa Rita, na cidade de Pelotas/RS nas na questão do consumo e geração de lixo, principalmente de copos e garrafas plásticas. O projeto contou com 62 alunos voluntários e 415 entrevistas.	pessoas daquela comunidade bebem água no seu cotidiano, assim como o registro dos dados na plataforma Fast Science.
Projeto Cidadão Cientista			
Ativo. Início em 2014.	Projeto desenvolvido pela SAVE Brasil, tem como parceiros a Divisão Técnica de Medicina Veterinária e Manejo da Fauna Silvestre (DEPAVE-3) e o Observatório de Aves do Instituto Butantan, no Brasil, e a Cornell Lab of Ornithology, dos EUA. É apoiado pela Fundação Grupo Boticário de Proteção à Natureza	O projeto foi desenvolvido pelo Instituto Tartabinhas, tendo como proposta o monitoramento subaquático de tartarugas marinhas através da fotoidentificação, na região de Bombinhas/SC, assim como a utilização das informações nas ações de conservação, sensibilização da população e criação de políticas públicas para a proteção dos ecossistemas marinhos e terrestre. Nos avistamentos novos, o voluntário nomeia simbolicamente a tartaruga e recebe um certificado simbólico da nomeação (Instituto Tartabinhas).	A participação dos voluntários ocorre através da interação de fotos, onde o voluntário pode fotografar os dois lados da cabeça da tartaruga e enviar para a base de dados para ser analisada.

QUADRO 1. PROJETOS DO SIBBR NO BRASIL E ANÁLISE DAS CARACTERÍSTICAS

Análise dos resultados

Com base no resultado das buscas no SiBBR, foram analisados os projetos em biodiversidade, assim como a forma como os não cientistas interagem com a ciência cidadã por meio da participação e seu desempenho nos projetos.

Os projetos de ciência cidadã que envolvem a biodiversidade, destacam a importância do engajamento dos voluntários não cientistas, pois graças a ajuda deles os cientistas que trabalham com meio ambiente podem ter acesso a uma grande quantidade de dados sobre a variedade de seres vivos presentes na natureza em curto espaço de tempo e amplo espaço geográfico, visto que muitos projetos abarcam todo o território nacional.

Nos projetos analisados é possível perceber que os de grande volume e visibilidade estão centrados nas regiões Norte, Nordeste, Centro-Oeste e Sudeste do país, sendo que na região Sul, verifica-se que são menores e se centralizam em pequenos nichos, mas com grande relevância para a sociedade e biodiversidade local.

Outro destaque relevante dos projetos é quanto à utilização das tecnologias para a coleta dos dados, seja através de aplicativos desenvolvidos para esse fim, plataformas ou até mesmo um servidor voltado para o armazenamento dessas informações.

Por fim, é visível a diversidade dos assuntos abordados nos projetos desenvolvidos através da Ciência Cidadã em biodiversidade no Brasil, além do compromisso tanto dos pesquisadores quanto dos voluntários em busca de resultados positivos por meio da conservação e preservação da nossa biodiversidade, aspecto importante para a sustentabilidade de uma região ou país.

Considerações finais

O estudo realizado buscou evidenciar a importância da Ciência Cidadã em projetos de biodiversidade desenvolvidos no Brasil. Foi possível identificar por

meio da plataforma SiBBR, alguns projetos relevantes para a área, atingindo os objetivos propostos.

No decorrer do trabalho, foi constatada a crescente preocupação por parte da sociedade, em relação à biodiversidade e à sustentabilidade do meio ambiente. Essa preocupação é verificada no envolvimento e contribuições dos voluntários nesses projetos, bem como nos resultados finais.

Além disso, pode-se perceber que alguns projetos não disponibilizam dados como o número de voluntários envolvidos, registros de observações realizadas e/ou espécies avistadas, fazendo com que não haja transparência de informações importantes, refletindo nos objetivos finais de cada projeto abordado.

Apesar da elevada tendência de crescimento de projetos voltados à Ciência Cidadã e biodiversidade no Brasil, e todo o engajamento entre pesquisadores e colaboradores voluntários, conclui-se que ainda há muito a ser explorado quanto à transparência das informações coletadas, pois os projetos são incipientes frente a todo um emaranhado de informações e dados a serem investigados, deixando algumas lacunas quanto ao preparo da Ciência Cidadã diante das mudanças na preservação e conservação da biodiversidade brasileira.

Bibliografia

ALVAREZ, G. R. y CAREGNATO, S. E. (2017) Preprints na comunicação científica da Física de Altas Energias: análise das submissões no repositório arXiv. *Perspectivas em Ciência da Informação*, 22(2). <https://doi.org/10.1590/1981-5344/2830>

AUDUBON. History of the Christmas Bird Count. <https://www.audubon.org/conservation/history-christmas-bird-count>

BRASIL. (2020). A Convenção sobre Diversidade Biológica. Ministério do Meio Ambiente. <https://www.gov.br/mma/pt-br/assuntos/biodiversidade/convencao-sobre-diversidade-biologica>

- CASTRO MARTINS, H. (2020). A importância da Ciência Aberta (Open Science) na pesquisa em Administração. *Revista de Administração Contemporânea*, 24(1). <https://doi.org/10.1590/1982-7849rac2020190380>
- MCGRATH, M. (2019). 1 milhão de espécies ameaçadas: o que diz preocupante relatório da ONU sobre impacto humano. <https://www.bbc.com/portuguese/geral-48174634>
- MENEGASSI, D. (2021). Sete aves da Mata Atlântica foram extintas nas últimas décadas. OECO. <https://www.oeco.org.br/reportagens/sete-aves-da-mata-atlantica-foram-extintas-nas-ultimas-decadas/>
- NASCIMENTO, A. P. (2018). Status atual e possibilidades da Ciência Cidadã para a conservação da biodiversidade no Brasil. (Trabalho de Conclusão de Curso, Universidade Federal do ABC). https://ccnh.ufabc.edu.br/arquivos/CENTRAL/4.Ensino/1.Graduacao/TCCs/TCC_Amanda-Porto_corrigido_final.pdf
- NAÇÕES UNIDAS (2021). Objetivos de Desenvolvimento Sustentável (ODS). <https://brasil.un.org/pt-br/sdgs>
- PARRA, H. Z. M. (2015). Ciência Cidadã: modos de participação e ativismo informacional. In: Albagli, S., Maciel, M. L. & Abdo, A. H. (Org.). *Ciência aberta, questões abertas*. (pp. 121-141). Ibict; Unirio. https://www.academia.edu/16137719/Ci%C3%Aancia_Cidad%C3%A3_modos_de_participa%C3%A7%C3%A3o_e_ativismo_informacional
- SISTEMA DA INFORMAÇÃO SOBRE A BIODIVERSIDADE BRASILEIRA (SiBBr). O que é Ciência Cidadã? <https://sibbr.gov.br/cienciacidadada/projetos.html>
- TEIXEIRA, L. A., COSTA, L.T. y ZAHER, E. H. (2015). Vital Brazil: um pioneiro na prática da Ciência Cidadã. <https://bibliotecadigital.butantan.gov.br/arquivos/38/PDF/3.pdf>
- VILLAR, R. R. (2020). O que é biodiversidade? Entenda a importância de preservar a natureza. <https://www.greenpeace.org/brasil/blog/o-que-e-biodiversidade-entenda-a-importancia-de-preservar-a-natureza/>
- YAMAMOTO, E. (2020). Programa Ciência Cidadã premia oito projetos para aproximar a pesquisa da comunidade. <https://jornal.usp.br/institucional/oito-projetos-que-aproximaram-a-pesquisa-cientifica-da-comunidade/>

Eje 2: Datos abiertos

Colecciones públicas abiertas: propuesta de depósito digital de metadatos bibliográficos de las publicaciones oficiales de Bogotá

WILSON ENRIQUE COLMENARES MORENO

Instituto Caro y Cuervo, Observatorio Editorial Colombiano

wilson.colmenares@caroycuervo.gov.co

RESUMEN

Las publicaciones oficiales requieren de metadatos bibliográficos basados en estándares que definan los campos básicos legibles por los sistemas de información, proporcionen la sintaxis necesaria para la descripción que garantice su recuperación precisa, interoperabilidad e intercambio de datos abiertos. La investigación tuvo como objetivo caracterizar los procesos de edición, organización, conservación y acceso a las publicaciones oficiales de la administración pública distrital y proponer lineamientos para la creación del depósito digital de metadatos a través de la Biblioteca Digital de Bogotá y contribuir con la consolidación del modelo de gestión del conocimiento y gobierno abierto de la ciudad. La metodología utilizada se desarrolló en siete fases, se realizó una entrevista y aplicaron dos encuestas en línea a 46 entidades. Los resultados permitieron identificar que Bogotá es la ciudad que más publicaciones oficiales produce del país con más de 500 títulos al año; el tamaño de las colecciones de trece de las bibliotecas es de 343.122 títulos en soporte papel y de 71.903 en soporte digital, sin embargo, se identificó que de las 44 entidades sólo ocho aplican algún estándar de metadatos, únicamente ocho tienen un repositorio institucional y solamente una entidad puede entregar los metadatos en un formato interoperable.

PALABRAS CLAVE

Edición de textos; metadatos bibliográficos; sistemas de información; procesamiento de la información. Bibliographic metadata; information systems; information processing; publishing.

Introducción

Recordar la metáfora de Max Weber de hace casi un siglo, en la cual hablaba del acceso a la información oficial a través de una ventana, es hoy real gracias a la web. Actualmente es posible asegurar que la web ha abierto una gran ventana de información a todos los ciudadanos. La llamada sociedad de la información ha provocado, fundamentalmente por la creación de internet y la World Wide Web, un cambio en la función de la administración de las instituciones públicas y en la actitud de los ciudadanos frente a su gestión. La web se ha convertido en el principal y global medio de comunicación de información textual, brinda la posibilidad de crear nuevas formas de trabajo que facilitan la creación de recursos de información más útiles para la sociedad.

Los constantes cambios en la organización del conocimiento, la creciente desarticulación de las estructuras institucionales y de los sistemas de información, están poniendo en relieve la importancia de los metadatos para la identificación, descripción y articulación del conocimiento producido por las sociedades. El acceso a la información por la mayor cantidad de personas desde cualquier lugar y a través de diferentes dispositivos tecnológicos es uno de los objetivos de las políticas de información a nivel mundial. Los países democráticos tienden a orientar sus sistemas de información hacia una sociedad de la información global. Este hecho promueve inquietudes, tendencias y un sinfín de perspectivas de futuro, sobre todo en el ámbito de los estándares y de la interoperabilidad entre los diferentes sistemas informáticos.

En la llamada sociedad informacional, CASTELLS enfatizó que:

(...) la fuente de la productividad estriba en la tecnología de la generación del conocimiento, el procesamiento de la información y la comunicación de símbolos. Sin duda, el conocimiento y la información son elementos decisivos en todos los modos de desarrollo, ya que el proceso de producción siempre se basa sobre cierto grado de conocimiento y en el procesamiento de la información. (1999, p. 43)

Sin embargo, entre más información produce la humanidad se hace cada vez más evidente la falta de normalización y aplicación de estándares internacionales como una alternativa para solucionar el problema de recuperación de la información en la web, pero también como un modelo conceptual y técnico que permita fortalecer la organización y la recopilación de la producción editorial. Así pues, el concepto de *información* en las mismas palabras de CASTELLS (1999, p. 43) es definido como «La información son los datos que se han organizado y comunicado». Y el *dato* es la unidad mínima de información con la posibilidad de representar hechos, conceptos o instrucciones y con la cualidad de ser comunicado, interpretado o procesado por humanos o sistemas informáticos. Ahora bien, los *datos abiertos* son datos digitales puestos a disposición del público con las características jurídicas y técnicas necesarias para que puedan usarse, reutilizarse y redistribuirse libremente por cualquier persona, en cualquier momento y en cualquier lugar (FUENTE KUNIGAMI & SERALE, 2018, p. 11). En esta medida, para lograr un uso más significativo, eficiente e integrador basado en las relaciones y redes que se pueden construir a partir de los objetos editoriales pensados como un objeto de información, se hace necesario abordar el concepto de *datos abiertos enlazados*, el cual va más allá de los datos abiertos, se puede entender como un método de publicación de datos estructurados a través de principios, como por ejemplo los principios FAIR (encontrables, accesibles, interoperables y reutilizables del inglés *Findable, Accessible, Interoperable and Reusable*)¹ y de

¹ Principios rectores FAIR, acrónimo del inglés que indica que los datos son encontrables, accesibles, interoperables y reutilizables. Se recomienda consultar la Iniciativa de datos abiertos del Gobierno de España: <https://datos.gob.es/es/noticia/principios-fair-buenas-practicas-para-la-gestion-y-administracion-de-datos-cientificos>

tecnologías web estándar para que los datos abiertos puedan ser vinculados, organizados y visualizados para diferentes propósitos.

Ahora veamos el término de *metadatos* enfocado a los recursos de información, en palabras de MÉNDEZ (2002), hace referencia a cualquier dato que ayude en la identificación, descripción y localización de un recurso de información. Son, en esencia, datos que describen los atributos de un recurso, caracterizan sus relaciones, permiten su recuperación, gestión y uso efectivo. Por ejemplo, los metadatos proporcionan información básica como el autor, año de creación, características de forma, palabras clave, identificadores persistentes, referencias y citas a otras publicaciones, etc. Además de estos metadatos básicos, cada vez cobra mayor importancia la descripción analítica del contenido a través de vocabularios especializados según sea el contenido, formato, audiencia y región geográfica —por ejemplo, género literario, tema, materia— para que los metabuscadores puedan encontrar y recopilar los metadatos que permitan vincular y visualizar los libros, las revistas, los periódicos y otras formas editoriales impresas o digitales, en diferentes catálogos y sistemas informáticos web.

La información oficial es toda aquella información emitida por cualquier organismo de la administración pública en el uso de sus funciones, sin importar el tipo de soporte documental y puede ser de reserva legal; por otra parte, las publicaciones oficiales son una forma editorial que ha sido financiada, avalada y producida por una institución oficial o en encargo por ella, su contenido es emanado de la autoridad del Estado según sus funciones, tienen autenticidad y son editadas en cualquier tipo de soporte o formato con la finalidad de ser comunicadas de forma masiva a la ciudadanía garantizando su integridad y normalización editorial (COLMENARES, 2015, p. 31). La información que se materializa en las publicaciones oficiales afecta la vida de cada ciudadano debido a su origen gubernamental, lo que las convierte en fuente de autoridad. Por consiguiente, la información que contienen las publicaciones oficiales no puede obtenerse de ninguna otra fuente, lo que las hace ser únicas e importantes para la sociedad. Un aspecto fundamental en el derecho a la información es el interés por parte de los ciudadanos de conocer las

actividades y las decisiones tomadas por los gobiernos; dicho interés provocó que los Estados iniciaran estrategias de edición y publicación de la información que producían, principalmente debido a que la información comenzó a considerarse como un derecho democrático básico en las sociedades. En consecuencia, las publicaciones oficiales surgieron con la materialización del proceso de edición y comunicación establecido entre la administración y los ciudadanos; este proceso se constituye en un imperativo al momento de garantizar la transparencia y la participación ciudadana en las acciones gubernamentales.

Por otra parte, el concepto de *depósito digital de metadatos bibliográficos* es pensado desde la importancia y necesidad de crear planes que definan lineamientos, procesos, infraestructura e instituciones responsables de fomentar la producción descentralizada, conservación, interoperabilidad, recopilación, curaduría y visualización de metadatos bibliográficos. En esta perspectiva el control bibliográfico como técnica histórica creada por las bibliotecas es fundamental, entendida como «un programa mediante el cual un país vela porque al menos un ejemplar de cada publicación se conserve para que cualquier investigador de cualquier parte del mundo que necesite consultarlo puede obtenerlo mediante préstamo o copia» (UNESCO, 2000, p. 9). Este tipo de programas han sido implementados en los diferentes países, por ejemplo, en Colombia es definido el depósito legal como la:

obligación que se le impone a todo editor de obras impresas, productor de obras audiovisuales y productor de fonogramas en Colombia y a todo importador de obras impresas, obras audiovisuales y fonogramas, de entregar para su conservación en las entidades y por las cantidades determinadas en el presente decreto, ejemplares de la obra impresa, audiovisual o fonograma producidos en el país o importados, con el propósito de guardar memoria de la producción literaria, audiovisual y fonográfica y acrecentar el patrimonio cultural. (Decreto 1.080 de 2015, art. 2.8.1.1)

Bajo este enfoque son fundamentales los estándares de producción, descripción, conservación e interoperabilidad de metadatos bibliográficos a

través de internet, ya que son la base para la creación de un programa de depósito digital de metadatos bibliográficos de las publicaciones oficiales. Aquí es importante mencionar el trabajo realizado por GILLILAND (2008), ZENG y QIN (2008) en el que clasifican los metadatos en cuatro tipos de estándares: 1) Metadatos de estructura: definen los campos básicos legibles por los sistemas de información. 2) Metadatos de contenido: proporcionan la sintaxis necesaria para la descripción de los metadatos. 3) Metadatos de valor: son los términos, nombres y otros valores que complementan la estructura de los datos o conjuntos de elementos de los metadatos y 4) Metadatos de intercambio: permiten la codificación y marcado para el procesamiento y transferencia de datos entre diferentes sistemas de información.

La necesidad de sustentar la toma de decisiones en evidencias, investigar los fenómenos relacionados con la administración pública y de profundizar en los hechos históricos, unida a la exigencia de transparencia administrativa y participación ciudadana demandan la creación de un nuevo modelo de gestión y acceso a los metadatos de las publicaciones oficiales sustentado en las tecnologías de la información y la comunicación, específicamente en la web. En la mayoría de Estados democráticos hay un decidido propósito de expandir el uso de las TIC para garantizar los derechos de los ciudadanos, teniendo en cuenta que el acceso a la información oficial no es un asunto únicamente técnico o tecnológico, sino de gobernanza democrática.

Las instituciones oficiales están en la obligación de dar a conocer a los ciudadanos la información en la cual soportan sus decisiones, como medio de transparencia de sus actos y con el fin de potenciar la confianza de la sociedad en sus instituciones oficiales. Esto conduce a que la información oficial sea de gran interés para las mismas administraciones públicas que la producen, para las empresas privadas y por supuesto para los ciudadanos en general. Parafraseando a STENSON (2005), quien concluía en su tesis doctoral que muchas cosas hacen pensar que una adecuada producción, edición, publicación, organización y acceso a las publicaciones oficiales supone una clara mejora organizacional, así no estén bien definidas ni las causas ni el alcance de dicha mejoría. Por lo tanto, una adecuada gestión de las

publicaciones oficiales es decisiva, tanto por el valor en sí mismas como por los activos económicos que representan.

Países europeos como España han centralizado el acceso a sus publicaciones oficiales desde la web², permitiendo acceder en tiempo real al texto completo de las publicaciones oficiales del Estado. Es importante destacar que a partir del año 2009 la publicación y el acceso desde la web tienen carácter oficial y auténtico. Para que el acceso en línea a las publicaciones oficiales del Gobierno Español fuera una realidad tuvieron que actualizar la legislación existente a la realidad de los avances tecnológicos del momento. La legislación española en materia de acceso abierto a sus publicaciones oficiales es un ejemplo que ilustra la posibilidad de brindar acceso a través de internet a la información oficial. Algo semejante ocurre en Alemania: el gobierno alemán brinda acceso a sus publicaciones oficiales a través de la web³, publica y almacena mensualmente de dos a tres millones de publicaciones, noticias, informes y otros documentos que a su vez reciben más de 25 millones de consultas. Dichas consultas muestran que la información oficial no solo es de interés para los funcionarios o profesionales relacionados con los temas gubernamentales, es de interés del público en general. Asimismo, el gobierno francés a partir de 2004 publica en internet⁴ sus publicaciones oficiales con el mismo valor legal que se le venía confiriendo a la publicación impresa. Y en Italia la principal publicación oficial legislativa se puede acceder a través de un sitio web⁵ de forma gratuita al resumen de todas las series publicadas desde 1988.

Una experiencia más cercana a nuestra región es la de Estados Unidos: en este país los estudios relacionados con las publicaciones oficiales tienen gran importancia debido a que existe un sólido sistema de edición, publicación y acceso web⁶. Tómese como ejemplo el caso de la Imprenta Oficial del Gobierno

² <http://www.boe.es> [Consulta: 23/06/2021].

³ <https://www.bgbl.de> [Consulta: 23/06/2021].

⁴ www.journal-officiel.gouv.fr [Consulta: 23/06/2021].

⁵ <http://www.gazzettaufficiale.it> [Consulta: 23/06/2021].

⁶ <http://www.gpo.gov> [Consulta: 23/06/2021].

Federal de los Estados Unidos (GPO): dicha institución tiene la función principal de imprimir y distribuir las publicaciones de toda la comunidad federal estadounidense, es así que el Congreso, la Casa Blanca y unos 130 departamentos federales y agencias gubernamentales deben enviar sus obras y comunicaciones a esta oficina de publicaciones para que puedan editarse con el objetivo de mantener informada a la administración pública y a los ciudadanos en general. La GPO elabora repertorios, catálogos y bases de datos con las que divulga las publicaciones disponibles y al mismo tiempo hace el control bibliográfico de las publicaciones.

En Colombia las publicaciones oficiales han sido analizadas únicamente por ROJAS, VÉLEZ y ECHEVERRI (1984), ARIAS (2007) y COLMENARES (2014), quien identificó que en el país no existe un catálogo colectivo o bibliografía de las publicaciones oficiales de la administración pública, lo que afecta profundamente la posibilidad de contar con un inventario de los activos de información del sector oficial como un primer paso para su identificación, control y utilización real por parte del Estado, de la empresa privada y de los ciudadanos. También identificó que

La administración pública de Bogotá no está en la capacidad de garantizar el acceso abierto a las publicaciones oficiales que producen sus instituciones oficiales debido a que no cuenta con una infraestructura tecnológica basada en estándares de preservación, descripción, acceso e interoperabilidad de metadatos. (2014, p. 196)

incluso, también estableció (como se presenta en la FIGURA 1) que la administración pública de Bogotá invierte altos recursos en la producción de estudios para soportar la toma de decisión, pero infortunadamente la mayoría de estudios no son editados y publicados oficialmente, lo que afecta el principio de publicidad de la información oficial, el cual constituye uno de los elementos característicos y esenciales del Estado Social de Derecho.

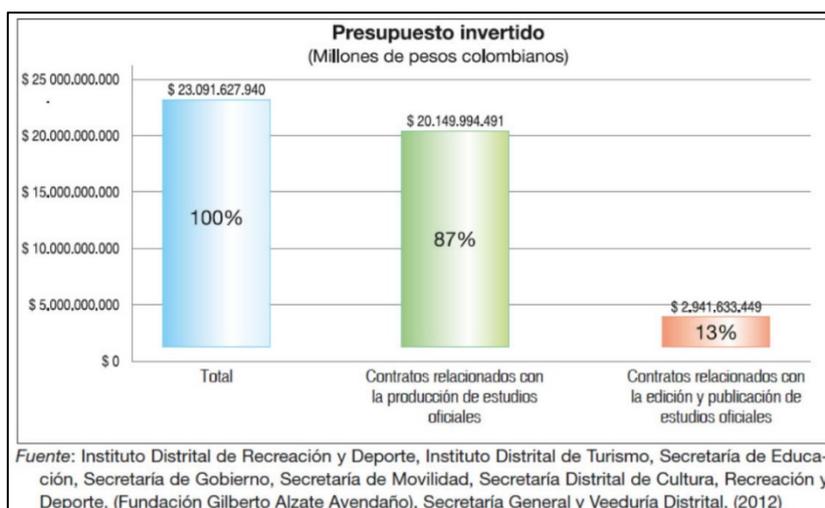


FIGURA 1. PRESUPUESTO INVERTIDO POR OCHO INSTITUCIONES GUBERNAMENTALES EN LA PRODUCCIÓN Y PUBLICACIÓN DE INVESTIGACIONES Y ESTUDIOS TÉCNICOS OFICIALES DE BOGOTÁ 2008-2012

Fuente: Colmenares, 2015, p. 31.

El anterior panorama hace evidente que la información pública se ha convertido en un gran recurso en la denominada sociedad de la información y para dar acceso y gestionar dicho recurso, la web y los sistemas informáticos han configurado una nueva realidad en relación con el acceso a la información producida por los Estados democráticos. Al respecto, GARCÍA (2008, p. 42) presenta algunos de los aspectos que confieren un valor agregado a la edición digital de publicaciones oficiales: 1) disminución de los costos de publicación; 2) simultaneidad en la edición y la publicación; 3) variedad de formatos de información; 4) mayor rapidez en la distribución y el acceso; 5) información actualizada; 6) eliminación de barreras geográficas y 7) permite elaborar otros subproductos de información.

Descripción del problema

El aumento de la edición y publicación en soportes y formatos digitales debido a la masificación de las tecnologías de información, relacionadas sobre todo con los bajos costos de edición, difusión y circulación de la publicación digital —entendida como un objeto de información— cuenta con varias problemáticas vinculadas con su fragilidad, compatibilidad, conservación, estabilidad e integridad digital y la necesidad de un mantenimiento constante

debido a la escalabilidad de los formatos. Estas características hacen de las publicaciones oficiales digitales un fenómeno mucho más complejo de organizar y gestionar, convirtiéndolas en un reto sustancial. Como se verá en los resultados de la presente investigación, la ciudadanía y la misma administración pública de Bogotá tienen diferentes tipos de problemas al momento de acceder de forma abierta desde la web a las publicaciones oficiales producidas por sus instituciones oficiales, principalmente para soportar los procesos de planeación, seguimiento y evaluación de las decisiones gubernamentales. Al respecto, cabe preguntarse ¿cuáles son los procesos de editoriales, organización, conservación y acceso a las publicaciones oficiales? ¿Son eficientes y eficaces los procesos de edición, organización, conservación y acceso a las publicaciones oficiales? ¿El derecho de autor es un impedimento para dar acceso abierto a las publicaciones oficiales? ¿La administración pública aplica estándares para dar acceso a las publicaciones oficiales y los metadatos bibliográficos? ¿Es posible hablar de gestión del conocimiento en la administración pública y de gobierno abierto sin brindar acceso abierto a los metadatos de las publicaciones oficiales?

Objetivo

La investigación tuvo como objetivo caracterizar los procesos de edición, organización, conservación y acceso a las publicaciones oficiales y a partir de los resultados obtenidos proponer lineamientos y recomendaciones para la creación del depósito digital de metadatos bibliográficos de la administración pública distrital a través de la Biblioteca Digital de Bogotá y contribuir con el fortalecimiento de la apropiación social del conocimiento institucional y la consolidación del modelo de gestión del conocimiento y de gobierno abierto de la ciudad.

Metodología

Se utilizó una metodología de investigación cualitativa de diseño transversal y de alcance exploratorio y descriptivo. Se produjeron datos primarios a partir de la aplicación de tres instrumentos de recopilación de información. El estudio se realizó entre abril y diciembre de 2020 a través de las siguientes fases: 1) Búsqueda, recopilación y sistematización de conceptos, estándares, lineamientos y metodologías relacionadas con las publicaciones oficiales y con sus principales esquemas de metadatos bibliográficos. 2) Diagnóstico de la producción, almacenamiento, recopilación, estandarización y acceso abierto a los metadatos de las publicaciones oficiales de Bogotá, para lo cual se aplicó una entrevista a las directivas de la Biblioteca Digital de Bogotá y un cuestionario de recopilación de información a la Imprenta Distrital de Bogotá, al Archivo Distrital de Bogotá, a la Cámara Colombiana del Libro y a la Biblioteca Nacional de Colombia. 3) Análisis de la legislación y la documentación vigente de los procesos de las secretarías del distrito relacionados con la edición, organización, preservación, comunicación y apropiación social de sus publicaciones oficiales. 4) Diseño y aplicación de una encuesta⁷ en línea con 57 preguntas enviada a través de correo electrónico a 77 entidades del Distrito y respondida por 44 entidades, entre ellas: secretarías, entidades adscritas, vinculadas, alcaldías locales y órganos de control. 5) Elaboración de la propuesta de depósito digital de metadatos bibliográficos de las publicaciones oficiales de Bogotá. 6) Creación de un directorio web de unidades de información de la administración pública de Bogotá y 7) Jornada de socialización de la propuesta de depósito digital de metadatos bibliográficos de las publicaciones oficiales de Bogotá.

Resultados y discusión

Los principales resultados de la investigación fueron:

⁷ Resultados de la Encuesta de Caracterización:
<https://coleccionedigitales.biblored.gov.co/items/show/1447> [Consulta: 23/06/2021].

Caracterización de las actividades de edición, organización, conservación y acceso a las publicaciones oficiales

De las respuestas obtenidas de las 44 entidades distritales encuestadas, se destacan las siguientes características según las:

Actividades de edición y de divulgación

Las actividades de edición y publicación contribuyen con la divulgación y la lectura de los contenidos inéditos. De acuerdo con la Encuesta, 35 entidades realizan actividades de edición, impresión o publicación y para realizar dichas actividades 23 entidades indicaron que no utilizan ninguna herramienta editorial. Se encontró que, de las 44 entidades encuestadas, 24 de ellas no asignan una licencia de derecho de autor a las publicaciones oficiales que producen y 9 entidades no saben o no respondieron la pregunta, lo que puede indicar que la gran mayoría de las publicaciones oficiales del Distrito producidas por las entidades encuestadas no son de acceso abierto. Respecto al cumplimiento de la asignación de ISBN o ISSN a las publicaciones oficiales que producen, 21 entidades no asignan este identificador, lo que dificulta el control bibliográfico y la generación de estadísticas de la producción editorial del Distrito. El depósito legal garantiza la conservación, divulgación y acceso a la memoria institucional del Distrito; sin embargo, según la Encuesta, 29 entidades no cumplen con la ley de depósito legal de las publicaciones oficiales que producen. Así mismo, 27 entidades indicaron que no envían un ejemplar de las publicaciones oficiales que producen a una biblioteca o centro de documentación. Las entidades encuestadas indicaron que tienen 5.626 títulos de sus publicaciones oficiales disponibles para la consulta y descarga en texto completo desde sus sitios web. Aun así, 23 entidades no cuentan con un inventario actualizado de las publicaciones oficiales que han producido. Por otra parte, 35 entidades no cuentan con un procedimiento de apropiación social de las publicaciones oficiales que han producido, lo que supone que se realizan muy pocas actividades de gestión del conocimiento institucional a partir de la producción bibliográfica.

Actividades de organización, conservación y acceso

Las bibliotecas y centros de documentación gubernamentales son unidades de información creadas y financiadas por las entidades gubernamentales, sus colecciones están integradas –principalmente– por publicaciones oficiales y los servicios que suministran buscan satisfacer las necesidades de información de los funcionarios del gobierno y de la ciudadanía. Al respecto, de las 44 entidades encuestadas, 20 de ellas cuentan con una biblioteca o centro de documentación. La Biblioteca del Concejo de Bogotá es la biblioteca gubernamental más antigua del Distrito, fue creada en 1928 y la Biblioteca Especializada del Archivo de Bogotá es la unidad de información gubernamental más joven, fue creada en el año 2020.

Las publicaciones oficiales son una forma editorial publicada en cualquier tipo de soporte y formato que ha sido financiada, avalada y producida por una institución oficial o en encargo por ella, garantizando su integridad y normalización editorial. No es de reserva legal. Al respecto, únicamente 5 de las unidades de información encuestadas indicaron que su personal tenía capacitación especializada en publicaciones oficiales. El procedimiento de adquisición de publicaciones garantiza el desarrollo de las colecciones de las unidades de información; con respecto a este procedimiento, 16 de las unidades de información encuestadas indicaron que tienen un procedimiento de adquisición de las publicaciones oficiales que produce la entidad. Más de la mitad de las unidades de información encuestadas indicaron que no tienen colecciones especializadas en publicaciones oficiales de Bogotá. Además, 5 de ellas indicaron que sus colecciones son de acceso restringido.

Respecto a la producción y gestión de metadatos bibliográficos, de las 20 unidades de información encuestadas, 8 indicaron que aplican alguna estructura o estándar de metadatos para describir las publicaciones oficiales de Bogotá que conservan. Aunque únicamente una unidad de información cuenta con un procedimiento que le permite entregar a otras instituciones los metadatos estructurados en un formato interoperable. En esta línea, los sistemas integrados de gestión de bibliotecas son un tipo de software

especializado en la automatización de los procesos de adquisición, catalogación, inventario, circulación, consulta y generación de estadísticas de colecciones bibliográficas. La mayoría de las unidades de información encuestadas indicó tener sistema integrado de gestión de bibliotecas, siendo el más usado Koha, con un total de 7 entidades distritales. De las 20 unidades de información encuestadas, tres de ellas tienen software de código abierto DSpace y una de ellas el software de código abierto Omeka. El servicio de diseminación selectiva de información permite mantener informados y actualizados a los usuarios según su necesidad específica de información a partir de las publicaciones que produce o adquiere la institución. De las 20 unidades de información encuestadas únicamente una presta el servicio de diseminación selectiva de información. Finalmente, únicamente 4 unidades de información generan reportes con información estadística de las publicaciones oficiales de Bogotá que conservan.

Propuesta de depósito digital de metadatos bibliográficos

Según los resultados obtenidos en la caracterización realizada, es posible decir que Bogotá es la ciudad que más produce publicaciones oficiales del país en soporte papel y digital. En esta medida, algunas entidades distritales cuentan con diferentes sistemas informáticos a través de los cuales almacenan, gestionan y brindan acceso a los metadatos de las publicaciones oficiales que producen; sin embargo, la gran mayoría de dichos sistemas no están articulados, la información es registrada por diferentes personas, con distintas políticas, normas y procedimientos que no cumplen estándares. Ninguna institución gubernamental del nivel distrital o nacional recopila, verifica, complementa, conserva y brinda acceso de forma centralizada e interoperable a los metadatos de las publicaciones oficiales de Bogotá. Para identificar, preservar y divulgar el patrimonio bibliográfico de la administración pública de la ciudad, se hace necesario recopilar los metadatos de las publicaciones oficiales que producen las entidades distritales según las nuevas formas de producción editorial. En particular, la Red de Bibliotecas Públicas de

Bogotá (BibloRed), tiene entre sus funciones la conservación, la difusión y el acceso a la información pública y el patrimonio bibliográfico de la ciudad. Y la Biblioteca Digital de Bogotá –iniciativa a cargo de BibloRed– funciona bajo un modelo de recopilación de metadatos de los diferentes sistemas de información (catálogos bibliográficos, repositorios, bibliotecas digitales, sistema de publicación) de las entidades del distrito que cumplan los estándares de interoperabilidad OAI-PMH. Por esta razón, la presente propuesta integra un conjunto de lineamientos que tienen como objetivo crear el depósito digital de metadatos bibliográficos de las publicaciones oficiales de la administración distrital a través de la Biblioteca Digital de Bogotá; para fortalecer la conservación, la divulgación, el acceso y la apropiación social del conocimiento institucional del distrito y contribuir con el gobierno abierto, la transparencia, la cultura política y la gestión de lo público.

Los lineamientos están compuestos de un conjunto de criterios y del Índice de Capacidades Editoriales y de Adquisición, Descripción, Conservación y Acceso a los Metadatos Bibliográficos de las Publicaciones Oficiales. Las ideas fuerza para la incorporación de los lineamientos son:

- Fortalecer la protección del derecho de autor de las entidades del distrito.
- Racionalizar el gasto público en la contratación de estudios técnicos y en la edición y publicación editorial.
- Fortalecer la recuperación, identificación y valoración de las publicaciones oficiales de Bogotá.
- Fortalecer la eficiencia y colaboración en y entre las entidades del Distrito, así como con los ciudadanos.
- Contribuir con el incremento de la transparencia en la gestión pública facilitando el acceso a la información oficial y el cumplimiento de este derecho constitucional.

- Incrementar la cultura política y promover la participación ciudadana a partir del fortalecimiento de la eficiencia en la recuperación, acceso y consulta de la producción editorial del Distrito.

La definición de estos lineamientos implica identificar un conjunto de responsabilidades necesarias para la implementación de la propuesta de depósito digital de metadatos bibliográficos; según los lineamientos definidos por MINTIC (2011), se mencionan algunas de las responsabilidades: 1) Establecimiento de políticas de información pública. Hace referencia a la responsabilidad en la definición de los lineamientos de política para la gestión de la información pública y la gestión de las tecnologías de la información que la soporta. Así mismo corresponde a la articulación de las políticas de datos abiertos con las políticas definidas para la información del Estado. Igualmente, la iniciativa debe estar alineada con la política de Buen Gobierno y con aquellas que impulsen la transparencia, la participación y la colaboración. 2) Oferta y publicación de datos abiertos. 3) Adquisición y mantenimiento de plataforma tecnológica. 4) Concierne a las responsabilidades de adquisición, diseño, construcción, implementación, mantenimiento y soporte de las plataformas tecnológicas que sustenten la oferta y demanda de datos abiertos y de aplicaciones construidas con ellos. 5) Definición de lineamientos tecnológicos. Incluye las responsabilidades de definición de estándares tecnológicos vinculados y asociados con la oferta y demanda de datos abiertos. 6) Acompañamiento a entidades. 7) Promoción de cultura de datos abiertos. 8) Implementación, apropiación y seguimiento de políticas por sector. Corresponde a las actividades de liderazgo sectorial que permitan garantizar la implementación y apropiación de la propuesta en cada uno de los sectores del gobierno. 9) Monitoreo y evaluación. Incluye actividades de observación del cumplimiento en la implementación de los lineamientos y políticas definidas, así como del seguimiento y evaluación general del impacto de la implementación de los lineamientos.

Índice de capacidades

Se creó el Índice para el fortalecimiento de las capacidades editoriales, de adquisición, descripción, conservación y acceso a los metadatos bibliográficos de publicaciones oficiales de las entidades del Distrito. El Índice está compuesto por 13 criterios distribuidos en dos capacidades. Su finalidad es hacer seguimiento a cada uno de los criterios y brindar recomendaciones para el fortalecimiento de las buenas prácticas.

Convenciones		
Capacidad alta	>=91%	
Capacidad media	>=51%	
Capacidad baja	<=50%	
No.	Nombre de la entidad	% capacidad total
1	Cámara de Comercio de Bogotá	93%
2	Veeduría Distrital	92%
3	Fundación Gilberto Alzate Avendaño	83%
4	Instituto para la Investigación Educativa y el Desarrollo Pedagógico IDEP	83%
5	Instituto Distrital de las Artes, Idartes	83%
6	Orquesta Filarmónica de Bogotá, OFB	76%
7	Secretaría de Educación del Distrito	76%
8	Jardín Botánico de Bogotá José Celestino Mutis	76%
9	Instituto Distrital de Patrimonio Cultural, IDPC	76%
10	Secretaría Distrital de Integración Social	75%
11	Archivo de Bogotá	62%
12	Instituto Distrital de Recreación y Deporte, IDRD	61%
13	Red Distrital de Bibliotecas Públicas, BiblioRed	61%
14	Secretaría Jurídica Distrital	52%
15	Secretaría Distrital de Ambiente	46%
16	Concejo de Bogotá	38%
17	Contraloría de Bogotá D.C.	36%
18	Secretaría Distrital de Desarrollo Económico	36%
19	Corporación Bogotá Región Dinámica, Invest In Bogotá	36%
20	Instituto Distrital de la Participación y Acción Comunal, Idpac	32%

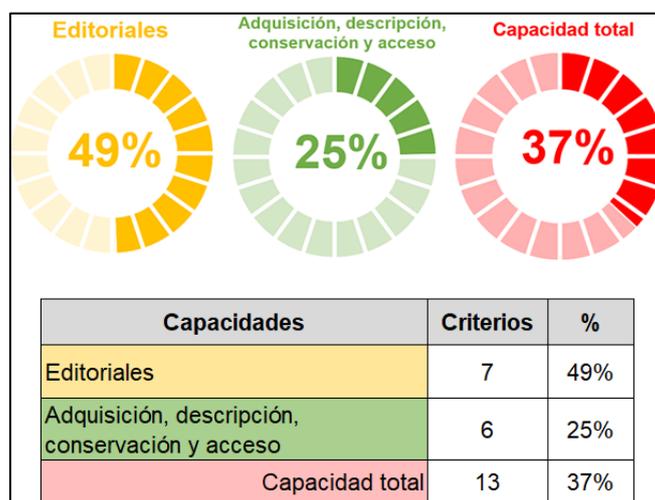


FIGURA 2. ÍNDICE PARA EL FORTALECIMIENTO DE LAS CAPACIDADES EDITORIALES, DE ADQUISICIÓN, DESCRIPCIÓN, CONSERVACIÓN Y ACCESO A LOS METADATOS BIBLIOGRÁFICOS DE PUBLICACIONES OFICIALES DE LAS ENTIDADES DEL DISTRITO

Elaboración propia

Lineamientos

Los lineamientos articulan lo dispuesto en la siguiente legislación relacionada con la impresión, depósito legal, acceso y conservación del patrimonio bibliográfico y documental del Distrito: Decreto Distrital 173 de 2004, Acuerdo Distrital 257 de 2006, Decreto Distrital 514 de 2006, Decreto Distrital 267 de 2007, Decreto Distrital 054 de 2008, Decreto Distrital 084 de 2008, Ley 1.379 de 2010, Ley 1.712 de 2014, Resolución nacional 3.564 de 2015, Decreto Distrital 828 de 2018 y Ley 1.915 de 2018.

Lineamiento 1: Capacidades editoriales

Producción editorial

Las actividades de edición y publicación contribuyen con la divulgación y la lectura de los contenidos inéditos.

¿La entidad realiza actividades editoriales?

Índice: de 44 entidades 35 de ellas realizan actividades editoriales.

Herramientas editoriales

¿La entidad utiliza herramientas editoriales?

Índice: de 44 entidades 23 de ellas utilizan herramientas editoriales.

Licencia de derecho de autor

¿La entidad asigna una licencia de derechos de autor a las publicaciones?

Índice: de 44 entidades 11 de ellas asignan una licencia de derecho de autor.

Código ISBN o ISSN

¿La entidad asigna un código ISBN o un código ISSN a las publicaciones?

Índice: de 44 entidades 23 de ellas asignan un código ISBN o un código ISSN.

Depósito legal

¿La entidad cumple con la ley de depósito legal?

Índice: de 44 entidades 15 de ellas cumplen con la ley de depósito legal.

Depósito impreso

¿La entidad envía un ejemplar de las publicaciones que produce a alguna unidad de información?

Índice: de 44 entidades 17 de ellas envían un ejemplar.

Depósito digital

¿La entidad registra las publicaciones que produce en algún sistema informático disponible en internet que permita el acceso abierto al texto completo?

Índice: de 44 entidades 28 de ellas registran las publicaciones.

Lineamiento 2: Capacidades de adquisición, descripción, conservación y acceso

Adquisición de publicaciones

¿La unidad de información tiene un procedimiento que le permita recibir las publicaciones que produce su entidad?

Índice: de 44 entidades 16 de ellas tienen un procedimiento de adquisición.

Nivel de descripción

¿La unidad de información aplica algún nivel de descripción (catalogación)?

Índice: de 44 entidades 14 de ellas aplica algún nivel de descripción.

Estructura o estándar de metadatos

¿La unidad de información aplica alguna estructura o estándar de metadatos?

Índice: de 44 entidades 8 de ellas aplica alguna estructura o estándar.

Sistema integrado de gestión bibliotecaria

¿La unidad de información tiene un sistema integrado de gestión bibliotecaria (software)?

Índice: de 44 entidades 18 de ellas tienen un sistema integrado.

Repositorio institucional

¿La unidad de información tiene un sistema de repositorio institucional (software)?

Índice: de 44 entidades 8 de ellas tienen un repositorio institucional.

Entrega de metadatos estructurados e interoperables

¿La unidad de información puede entregar los metadatos estructurados en un formato interoperable?

Índice: de 44 entidades 1 de ellas puede entregar los metadatos.

Modelo del depósito digital de metadatos bibliográficos

Con base en los lineamientos propuestos, se estructuró la propuesta de modelo de depósito digital de metadatos bibliográficos de las publicaciones oficiales de Bogotá: es un modelo colaborativo que integra a los productores, recolector y usuarios de los metadatos de las publicaciones:

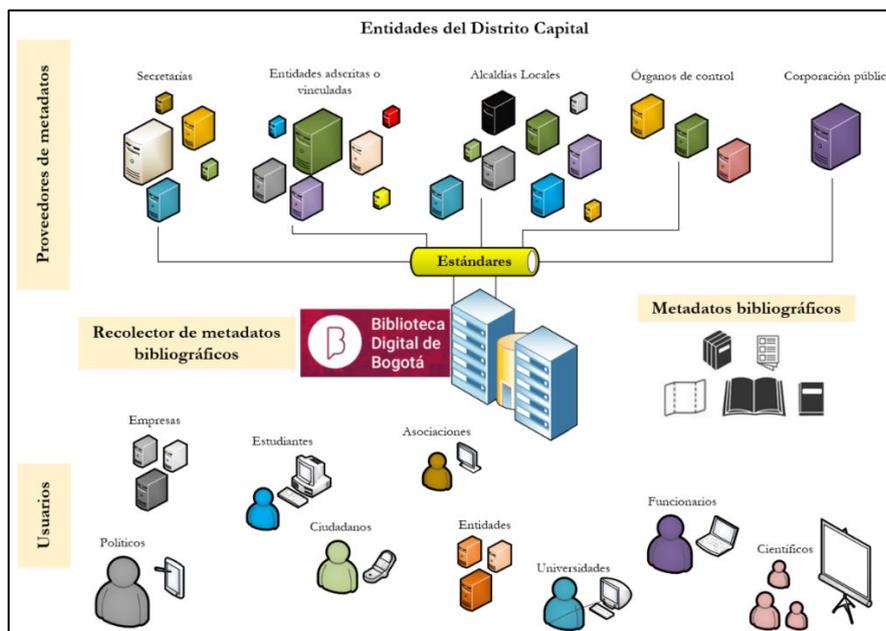


FIGURA 3. MODELO DEL DEPÓSITO DIGITAL DE METADATOS BIBLIOGRÁFICOS DE LAS PUBLICACIONES OFICIALES DE BOGOTÁ

Elaboración propia

Diagrama de flujo del procedimiento de depósito digital de metadatos

La producción de metadatos bibliográficos requiere de una ruta que describa las principales capacidades requeridas para la realización del depósito digital de metadatos.

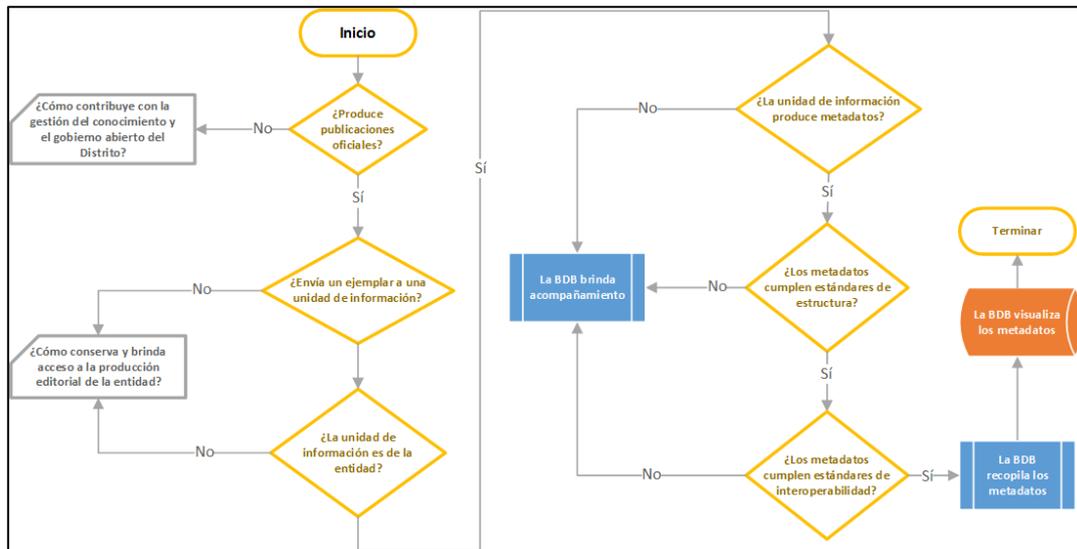


FIGURA 4. DIAGRAMA DE FLUJO DEL PROCEDIMIENTO DE DEPÓSITO DIGITAL DE METADATOS

Elaboración propia

Recursos y servicios derivados del depósito digital de metadatos bibliográficos

El depósito digital de metadatos bibliográficos de las publicaciones oficiales de Bogotá permitiría generar un conjunto de recursos y servicios de información que contribuiría con el gobierno abierto, la transparencia, la cultura política y la gestión de lo público.

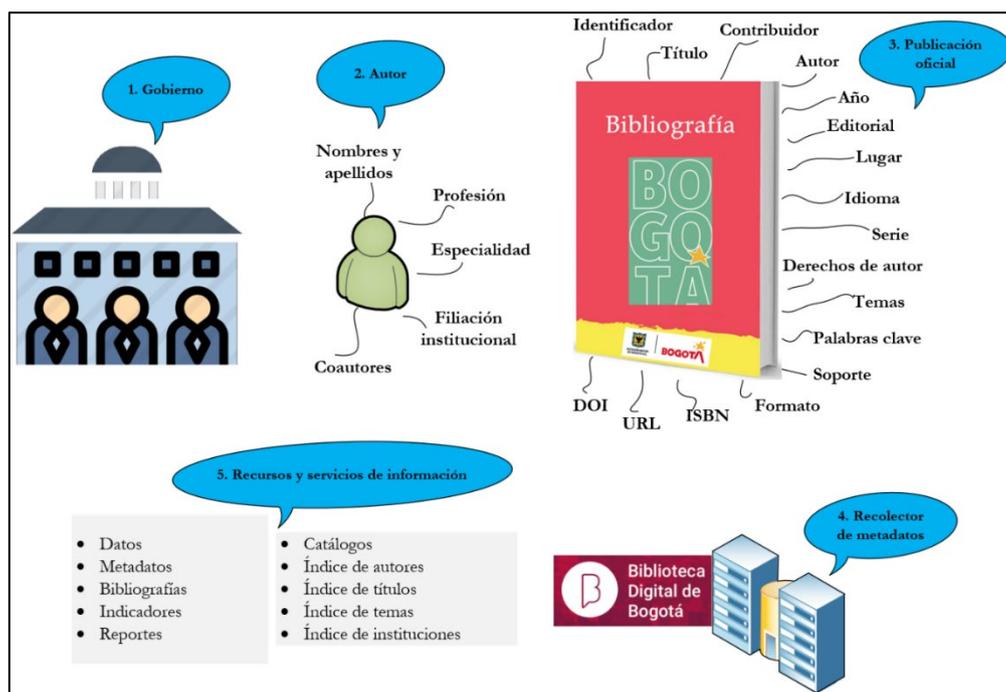


FIGURA 5. RECURSOS Y SERVICIOS DERIVADOS DEL DEPÓSITO DIGITAL DE METADATOS BIBLIOGRÁFICOS

Elaboración propia

Directorio de unidades de información de la administración pública de Bogotá

El Directorio es una herramienta web que permite identificar, acceder, registrar y compartir los datos de contacto y las trayectorias de las unidades de información de la administración pública de Bogotá. Su fin es fortalecer la gestión y la apropiación social del conocimiento institucional a través de la creación de la ruta de formalización administrativa del depósito digital de las publicaciones oficiales y contribuir con la implementación, de manera colaborativa, de un modelo de gobierno abierto para la ciudad. El Directorio fue creado en Wikidata, base de conocimientos editada en colaboración, proporcionando una fuente común de datos abiertos enlazados que pueden ser consultados, editados y reutilizados por cualquier persona, bajo una licencia de dominio público. El primer conjunto de datos fue recopilado a través de la Encuesta de Caracterización. A partir de los datos recopilados se crearon 35 declaraciones y se registraron más de 900 datos abiertos. Los perfiles de 18 unidades de información gubernamentales de Bogotá se encuentran publicados en [Wikidata](#).

Bibliografía

- ARIAS, J. (2007). Las publicaciones oficiales en Colombia: Situación actual y necesidades del control bibliográfico. *Bibliotecas & Tecnologías de la información*, 4(1), 34-44.
- CASTELLS, M. (1999). *La era de la información. Vol. I: economía, sociedad y cultura. La sociedad red*. Siglo XXI.
- COLMENARES, W. (2014). Diagnóstico del acceso abierto desde la web a las publicaciones oficiales de la administración pública de Bogotá [Trabajo de grado profesional en Sistemas de Información, Bibliotecología y Archivística, Universidad de La Salle. Facultad de Ciencias Económicas y Sociales. Programa Sistemas de Información, Bibliotecología y Archivística]. https://ciencia.lasalle.edu.co/sistemas_informacion_documentacion/28/
- COLMENARES, W. (2015). Las publicaciones oficiales de la administración pública de Bogotá, Colombia. En *La información gubernamental y el acceso a la información pública* (pp. 15-38). Universidad Nacional Autónoma de México, UNAM. Instituto de Investigaciones Bibliotecológicas y de la Información, IIBI. <http://eprints.rclis.org/31555/>
- GARCÍA SÁNCHEZ, M. I. (2008). Las publicaciones oficiales. Estado actual de la cuestión. *Revista General de Información y Documentación*, 18, 39-79.
- GILLILAND, A. J. (2008). Introduction to Metadata. Version 2.0. Setting the Stage. http://www.getty.edu/research/publications/electronic_publications/intrometadata/setting.html
- MÉNDEZ, E. (2002). *Metadatos y recuperación de información: Estándares, problemas y aplicabilidad en bibliotecas digitales*. Trea.
- Decreto 1080 de 2015. Por medio del cual se expide el Decreto Único Reglamentario del Sector Cultura, 1080 Decreto (2015). <https://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=62515>
- MINISTERIO DE TECNOLOGÍAS DE LA INFORMACIÓN Y COMUNICACIONES MINTIC. (2011). Lineamientos para la implementación de datos abiertos en Colombia. <https://www.academia.edu/29630319>
- MORTIMER, M. (2002). *Catalogación descriptiva* (2ª ed.). Rojas Eberhard.
- MUENTE KUNIGAMI, A. y SERALE, F. (2018). *Los datos abiertos en América Latina y el Caribe*. Banco Interamericano de Desarrollo. <http://gobiernoabierto.org.mx/documentos/libros/48.pdf>

- ORGANIZACIÓN DE LAS NACIONES UNIDAS PARA LA EDUCACIÓN, LA CIENCIA Y LA CULTURA. UNESCO. (2000). Legislación sobre depósito legal: Directrices. <http://unesdoc.unesco.org/images/0012/001214/121413s.pdf>
- ROJAS, O., VÉLEZ, V., y ECHEVERRI, R. (1984). Disponibilidad y uso de las publicaciones oficiales en Colombia. *Universitas Humanística*, 13(21), 21. <https://revistas.javeriana.edu.co/index.php/univhumanistica/article/view/10288>
- STENSON, J. (2005). *The attributes of information as an asset*. (Sin publicar. Consulta en sala.) [Tesis Doctoral]. Loughborough University.

Eje 3: Evaluación y métricas alternativas

Datos bibliométricos para las Ciencias Sociales y las Humanidades: un método para el acopio, validación y análisis con herramientas de acceso gratuito

SILVIA EUNICE GUTIÉRREZ DE LA TORRE

Biblioteca «Daniel Cosío Villegas», El Colegio de México

segutierrez@colmex.mx

JOSÉ VALENTÍN ORTIZ REYES

Biblioteca «Daniel Cosío Villegas», El Colegio de México

jvortiz@colmex.mx

JONATHAN ISRAEL ESCOBAR FARFÁN

Biblioteca «Daniel Cosío Villegas», El Colegio de México

jescobar@colmex.mx

TOMÁS BOCANEGRA ESQUEDA

Biblioteca «Daniel Cosío Villegas», El Colegio de México

tbocanegra@colmex.mx

VÍCTOR CID CARMONA

Biblioteca «Daniel Cosío Villegas», El Colegio de México

vjcid@colmex.mx

CLAUDIA ESCOBAR VALLARTA

Biblioteca «Daniel Cosío Villegas», El Colegio de México

cescobar@colmex.mx

MARÍA LOURDES QUIROA HERRERA

Biblioteca «Daniel Cosío Villegas», El Colegio de México

lourdes_verd@hotmail.com

CAMELIA ROMERO MILLÁN

Biblioteca «Daniel Cosío Villegas», El Colegio de México

cromero@colmex.mx

RESUMEN

La baja cobertura en los índices globales más importantes, como Web of Science (WOS) y Scopus, de la investigación en español de las Ciencias Sociales y las Humanidades (CSyH) obligan, a quienes analizan las citas que se hacen a la producción académica de sus comunidades, a explorar alternativas para recopilar datos de mayor alcance, confiables y validados. En este documento se describe un método que emplea Google Académico (GA) como principal fuente de información para el análisis de citación de los artículos publicados en revistas arbitradas por la planta académica de El Colegio de México entre los años 2012-2020. A partir de una base de datos compilada y validada desde 2016, por personal académico de la Biblioteca «Daniel Cosío Villegas»¹, se ofrece evidencia sobre el alcance de GA para el análisis de citación, destacando sus ventajas y limitaciones. Se describe el proceso de acopio, análisis e interpretación de los datos, el cual se realiza exclusivamente con software gratuito y de libre acceso (Google Drive, Zotero, R y GA). Con esto nos proponemos compartir un método que pueda ser replicado en contextos similares.

PALABRAS CLAVE

Análisis de citas; servicios bibliométricos; herramientas de acceso gratuito. Citation Analysis, Bibliometric Services, Free Access Tools.

¹ Tomás Bocanegra, Víctor Cid, Claudia Escobar, Israel Escobar, Silvia Gutiérrez, Valentín Ortiz, Carolina Palacios, Lourdes Quiroa, Camelia Romero, Arón Sánchez y en ediciones anteriores: Mariana Córdoba, Máximo Domínguez, Lourdes Guerrero, José Manuel Morales.

Introducción

En los últimos años el uso de métricas para valorar la producción académica ha tomado mayor relevancia. Por esta razón, la reflexión sobre las fuentes y métodos empleados para la obtención de datos se ha centrado en la búsqueda de parámetros más justos y apegados a las prácticas de publicación de cada disciplina. Para el caso del trabajo que se realiza en la Biblioteca «Daniel Cosío Villegas» (BDCV) de El Colegio de México (COLMEX), esta reflexión se ha centrado, por la naturaleza de los programas de formación y líneas de investigación que se desarrollan, en las Ciencias Sociales y las Humanidades (CSyH), tarea que derivó inicialmente del servicio de análisis de citas que se ofrece a la planta académica.

El COLMEX es una institución pública dedicada a la formación e investigación en CSyH, a nivel licenciatura y, principalmente, a nivel de posgrado. Cuenta con ocho centros de estudios: Centro de Estudios de Asia y África (CEAA), que cuenta con 26 personas en su planta académica; Centro de Estudios Demográficos, Urbanos y Ambientales (CEDUA), 30 personas; Centro de Estudios Económicos (CEE), 16; Centro de Estudios Históricos (CEH), 25; Centro de Estudios Internacionales (CEI), 24; Centro de Estudios Lingüísticos y Literarios (CELL), 30; Centro de Estudios Sociológicos (CES), 22 y el Programa de Estudios Interdisciplinarios (PEI), 50, que se integra por el personal académico jubilado.

Durante 2016 se analizaron los indicadores que ofrecen distintas fuentes de información sobre la citación de artículos publicados por la planta académica. A partir de aquel ejercicio se identificó que ninguna base de datos ofrece información confiable y actualizada para analizar las citas a nivel institucional. Al realizar el ejercicio de búsqueda con el filtro de filiación institucional en Scopus, se ubicaron 353 documentos publicados entre 2012 y 2016. Sin embargo, al revisar puntualmente los documentos, se ubicaron artículos cuyas temáticas estaban asociadas con ciencias duras. El error que identificamos fue que había autores que pertenecían a El Colegio Nacional, pero que para Scopus tenían filiación con el COLMEX. Además de esto, identificamos que los

números o indicadores que disponen las bases de datos están limitados al universo de documentos que albergan, es decir, no consideran otras fuentes de información para generar datos y los que ofrecen son limitados por su propio alcance.

Como adelantamos, los ejercicios de análisis de citas evidenciaron problemas de cobertura, pues el número de citas que se ubicaban en WOS o Scopus son mínimas, en comparación con las identificadas por medio de GA. La anotación al margen en este caso es que hablamos de textos relativos al estudio de las CSyH escritos en español, condiciones que restringen su representatividad en ambas bases de datos de pago (ver PLAZA *et al.*, 2018). En el análisis que realizamos en 2019, con la base de datos Scopus, sobre la distribución por idioma de artículos publicados, entre 2014 y 2018, en el área de Ciencias Sociales, encontramos que el 89 % fueron escritos en inglés y un 2,54 % en español. En un escenario similar, ubicamos en WOS, que el 87,12 % de los artículos publicados en el área de Ciencias Sociales fueron escritos en inglés y tan sólo un 3,11 % en español (*Las revistas científicas y el español*, 2019). Con todo, a pesar de ser la herramienta disponible más adecuada, está lejos de ser perfecta. En un análisis reciente se ha probado que la visibilidad de los artículos en español está significativamente afectada por los algoritmos de relevancia de GA, lo cual afecta la posibilidad de cosechar citas, pues a menor facilidad de encuentro, menor posibilidad de ser leído o citado (ROVIRA, CODINA, & LOPEZOSA, 2021).

Estudios como los de HARZING y ALAKANGAS (2016) han discutido la necesidad de realizar estudios longitudinales para tener mayor evidencia sobre la cobertura o alcance de Scopus, WOS y GA, pues el uso *per se* de estos recursos puede determinar las métricas o resultados obtenidos y, con ello, modificar las conclusiones que se puedan hacer sobre los datos. Parte de su argumentación apunta a que GA ofrece una cobertura más amplia, y, por lo tanto, indicadores más altos que otras fuentes de información. Sin embargo, requiere de un trabajo adicional de control de calidad, pues rastrea cualquier información que esté disponible en la red. Otro de los inconvenientes que apuntan estas autoras es el gran número de documentos duplicados, que

varían en detalles nimios de sus respectivos registros. El problema con el contenido que ofrece GA, de acuerdo con GINGRAS (2016) es que puede incluir documentos provenientes de publicaciones académicas, pero también toma documentos de páginas personales, documentos de trabajo, borradores, pruebas de imprenta. A esto se suma el riesgo, al que volveremos más adelante, de que los documentos citantes pueden aparecer o desaparecer, modificando así la validación de los indicadores, calculados por el propio recurso.

Para MARTÍN MARTÍN, ORDUNA MALEA, THELWALL y DELGADO LÓPEZ CÓZAR (2018), GA encuentra más citas que WOS Core Collection y Scopus en todas las áreas temáticas. A partir de un conjunto de datos de distintas disciplinas descubrieron que casi todas las citas encontradas por WOS (95 %) y Scopus (92 %) fueron también encontradas por GA. A esto se le suma que esta última base encontró una cantidad sustancial de citas únicas que no fueron ubicadas por las otras bases de datos. En las áreas de CSyH las citas únicas de GA, según este estudio, superan el 50 % de todas las citas del área.

Una de las herramientas que se valoró en 2016 para el acopio de información fue Publish or Perish, un programa de acceso gratuito desarrollado por Anne-Wil Harzing, que permite búsquedas avanzadas para el acopio y exportación de reportes numéricos de las citas provenientes de fuentes como GA y Microsoft Academic Research. Esta herramienta ha sido útil para otros reportes en los que no se hace la revisión manual de las citas pues, si bien la última versión permite recuperar los documentos citantes de cada ítem, este proceso requiere correr la búsqueda por cada artículo y hacer después el cotejo de la veracidad de la cita, tratamiento que describiremos más adelante.

Frente al requerimiento institucional de ofrecer evidencias sobre las citas a la producción académica del COLMEX, se delineó una estrategia para obtener información confiable sobre las citas a artículos publicados por la planta académica en revistas propias y ajenas, desde 2012 hasta el 2016. Estos son algunos de los elementos que se consideraron para dar forma a la base de datos que el personal académico, con el apoyo de personal auxiliar, ha actualizado y mantenido en los últimos cinco años y la cual ha servido como

principal insumo para la elaboración de reportes anuales². A continuación, se describe el método empleado para el acopio de datos, así como los criterios que se aplican en su validación. Se muestra en detalle el procedimiento de sistematización de los datos con el uso de herramientas como Google Drive, Zotero, R en RStudio y OpenRefine. Al final se exponen algunas consideraciones a tomar en cuenta en la aplicación de este método.

Descripción del método

Anualmente se analizan las citas reportadas por GA para los artículos publicados por profesores-investigadores del COLMEX en revistas académicas. En este análisis se buscan citas únicamente a los artículos que son reportados, por la propia comunidad académica, en el Currículum Electrónico COLMEX de cada profesor-investigador, y no se consideran las citas a libros, capítulos de libros y otros documentos. Estos registros, que se realizan de forma personal, no están exentos de errores de captura, pero para el estudio se normalizan los campos de año y lugar de publicación.

Se validan únicamente citas asentadas en publicaciones del periodo comprendido entre 2012 y el año inmediato anterior al del reporte anual. Por ejemplo, para el reporte del 2021, se analizaron sólo artículos que citan la producción COLMEX entre 2012 y 2020. Se excluyen las que se consideran citas «no válidas», sobre lo cual se ahondará más adelante. Para el proceso de validación se aplican los siguientes criterios de inclusión para seleccionar los trabajos citantes:

- Se validan sólo aquellos a los que se tiene acceso en texto completo y en cuya bibliografía se haya cotejado la cita al documento publicado por investigadores COLMEX.
- Se validan aquellos que no hayan sido escritos por la persona citada, es decir, que sean «autocitas».

² Estos informes están disponibles en: <https://www.colmex.mx/es/documentacion-institucional>

- Se consideran sólo publicaciones arbitradas. Así, se excluyen como citas válidas las hechas en: documentos de trabajo, ponencias, presentaciones y memorias de actos académicos publicadas sin ISBN o ISSN, o, en su defecto, que se encuentre en trámite. Si bien estos identificadores no aseguran que se trate de una publicación arbitrada, requieren un nivel de trámite superior a la simple subida de un archivo digital.
- Sólo se validan citas en tesis en las que la persona citada no haya tenido ninguna participación como tutor o lector.
- Se excluyen los documentos citantes publicados en el año en el que se realiza el análisis.

Procedimiento de acopio con Google Académico (GA) y Google Drive. En esta sección se describe la hoja de cálculo colaborativa, las búsquedas que realizan los compañeros auxiliares, y se detalla el uso que se da a las columnas de control.

La Bibliotecaria de Humanidades Digitales (BHD) crea en Google Drive una hoja de cálculo colaborativa para el año de análisis correspondiente, partiendo de una copia de la hoja de cálculo usada en el año anterior. Así se mantiene un registro histórico del análisis, y se aprovechan avances anteriores. En esta nueva hoja de cálculo se crea una pestaña para cada centro de estudios, y en cada pestaña se actualiza el listado de artículos. Cada artículo corresponde a un renglón de la hoja de cálculo y es descrito con las siguientes columnas: *Autor, Título del artículo, Nombre de la revista en la que fue publicado, Editorial, Lugar de publicación, Año, y Centro* (ver FIGURA 1).

	A	B	C	D	E	F	G
1	Autor	Título del artículo	Nombre de la revista	Editorial	Lugar de publicación	Año	Centro
2	Martín Butragueñ	A veces lloro mis lágrimas. Aproximac	Estudios de Lingüi	UNAM	México	2016	CELL
3	Martín Butragueñ	La concordancia de haber existencia	Boletín de Filologí	Universidad de Chile		2016	CELL
4	Martín Butragueñ	Aproximación al uso del modo subj	Boletín de Filologí	Universidad de Chile		2012	CELL
5	Pedro Martín Butr	Prosodia fonética de enunciados rep	Estudios de Fonét	Universidad de España		2014	CELL
6	Pedro Martín Butr	Hacia una prosodia basada en el usc	Normas	Universidad de España		2015	CELL
7	Luz Elena Gutiérre	La violenta transformación de la viol	Romance Notes 5	University of N	Estados Unic	2014	CELL
8	Sergio Eduardo Bc	Los clíticos pronominales del españ	Nueva Revista de	El Colegio de N	México	2015	CELL
9	Martín Butragueñ	Allá llega a lo que es el pueblo de Sa	Lingüística y Litera	Universidad de Colombia		2016	CELL

FIGURA 1. COLUMNAS PARA DESCRIBIR ARTÍCULOS EN GOOGLE DRIVE

Además de las columnas anteriores, se incluye una columna *Liga GA*, en la que se registra la URL de resultados para la búsqueda de cada artículo en GA. Cuando GA no reporta resultados para la búsqueda de un artículo de la lista, se consigna la nota: *No reconocido*. Colaboradores auxiliares son responsables de correr las búsquedas siguiendo la URL correspondiente, y actualizarla, si es el caso, para artículos no reconocidos en años anteriores, o para aquellos que recién se integran a la lista.

En cada búsqueda, se coteja el número de citas reportadas para el artículo, y se consigna en una columna *Citas GA XXXX*, donde XXXX corresponde al año analizado. En otra columna llamada *Diferencias* se calcula la diferencia entre las citas reportadas por GA el año anterior y las del año analizado. Por ejemplo, en el estudio de 2020 se comparan las citas reportadas por GA ese año, con las reportadas en el 2019. En principio, solamente aquellos artículos que hayan presentado una diferencia en citas reportadas son los que tendrían nuevas citas para validar.

A continuación, la BHD actualiza las columnas de control (ver FIGURA 2) donde XXXX corresponde al año de análisis, es decir, aquel en el que se realiza el ejercicio. Estas columnas sirven para distinguir el número de citas válidas, desglosando el número de citas «no válidas» reportadas por GA. Las *Citas Zotero* son las citas validadas de acuerdo con los criterios mencionados anteriormente. La columna *Sin Acceso* registra la cantidad de citas que no fue posible verificar porque no se tuvo acceso al texto completo del documento

citante. En *Autocita*, se cuentan aquellas publicaciones citantes cuyo autor es el mismo del artículo citado. *No arbitrado* sirve para contabilizar las citas hechas en documentos sin revisión, como se explicó previamente. En la columna *Otro* se cuentan documentos que no pertenecen a ninguna de las categorías anteriores, por ejemplo, duplicados, errores, tesis dirigidas por el profesor investigador citado, o publicaciones del año en el que se realiza el análisis.

										Columnas de control						
	E	F	G	I	J	K	L	M	N	O	P	Q	R	U		
1	Lugar de pul Año	Centro	Citas GA 2019	Citas GA 2020	Diferencias	Citas Zotero 2021	Sin acceso 2021	Autocita 2021	No arbitrado 2021	Otro 2021	Comprobacion	Liga GA	Notas			
2	México	2016 CELL	8	11	3	6	0	3	0	3	-1	https://scholar.google.com/scholar?cites=15556044187525	1 AC en He			
3	Chile	2016 CELL	11	12	1	6	1	2	0	3	0	https://scholar.google.com/scholar?hl=es&as_sdt=2005&sc	JIEF 2020: 2			
4	Chile	2012 CELL	21	25	4	19	0	3	1	2	0	https://scholar.google.com/scholar?cluster=452858605257	1 autocita e			
5	España	2014 CELL	15	22	7	7	2	9	0	4	0	https://scholar.google.com/scholar?cluster=250971789132	JIEF 2021: 2			
6	España	2015 CELL	7	10	3	8	0	2	0	1	-1	https://scholar.google.com/scholar?hl=es&as_sdt=0%2C5&sc	GS ya no lis			
7	Estados Unid	2014 CELL	4	7	3	4	1	0	0	2	0	https://scholar.google.com/scholar?hl=es&as_sdt=0%2C5&sc	JIEF 2021: 2			
8	México	2015 CELL	5	11	6	5	0	2	0	4	0	https://scholar.google.com/scholar?cluster=161426336412	JIEF 2021: 2			
9	Colombia	2016 CELL	7	11	4	8	0	3	0	1	-1	https://scholar.google.com/scholar?cites=13083579098513	JIEF 2021: 2			

FIGURA 2. COLUMNAS DE CONTROL EN GOOGLE DRIVE

Finalmente, la BHD integra una columna para *Comprobación* en la cual se calcula la diferencia entre la suma de las columnas de control y el número de posibles citas reportadas por GA para el año analizado. En principio, una diferencia de cero indicaría que se ha dado cuenta de todas las posibles citas reportadas por GA, y el número total de éstas se ha distribuido entre las columnas de control. Una columna para notas es utilizada para hacer las anotaciones que puedan ayudar, en ejercicios subsecuentes, a la mejor identificación de citas ya validadas o ya descartadas. Una vez preparada la hoja de cálculo colaborativa en Google Drive, se procede a la validación de citas y al registro de artículos citantes en Zotero.

Validación e integración de las citas en Zotero. En esta sección se describe la biblioteca compartida de Zotero, y se hacen precisiones sobre las citas a más de un artículo, y las coautorías. Los bibliógrafos asignados a cada centro de estudios son los encargados de validar las posibles citas reportadas por GA, y de compilar la información bibliográfica de los artículos citantes en una biblioteca compartida de Zotero

Como requisitos previos, cada bibliógrafo debe instalar un cliente de Zotero en su computadora de trabajo (STILLMAN, 2021). También debe instalar el complemento de Zotero para su navegador de internet, que le permite importar las referencias bibliográficas desde la lista de resultados de GA, o desde la base de datos donde se aloja el documento citante.

Para revisar las posibles citas a cada artículo, se pueden ordenar los datos en cada pestaña de centro por la columna *Diferencias*, para analizar aquellas que en principio tuvieron aumento de citas. Por cada artículo, se sigue la liga de GA con la lista de posibles documentos citantes, y de esta lista se validan las citas que cumplen los criterios establecidos, y se contabiliza cada cita «no válida» usando, en cada caso, las columnas de control descritas en la sección anterior.

Los datos bibliográficos de un documento citante que incluye una cita válida se integran a Zotero. Dicha biblioteca tiene la siguiente estructura (ver FIGURA 3). Esta estructura jerárquica permite concentrar la información por centro, autor y sus artículos citados. Las carpetas de cada artículo se nombran con una combinación del año de publicación del artículo y su título.

The screenshot displays the Zotero desktop application. On the left, a hierarchical tree view shows folders for 'BIBLIOTECA', 'CEEA', 'CEUA', 'CEE', 'CEH', 'CE', and 'CELI'. Under 'CELI', there are sub-folders for authors like 'Avila Sánchez, Francisco Faul' and 'Barriga Villanueva, Rebeca'. A folder named 'Hernández Rodríguez, Ernesto' is selected, containing a sub-folder '2020 Aprensión al uso del modo subjuntivo'. This folder contains several articles, with 'Prácticas y concepciones sobre la alternancia verbal escrita en pasado irreal al comentar un texto literario en bach...' highlighted.

The main window shows a table of bibliographic records. The selected record is highlighted in green:

Título	Creador	Tipo de elemento	Año
Degrees of subjunctive vitality among monolingual speakers of Peninsular and Argentinian Spanish	Gallego y Alonso-Marks	Artículo de revista académ...	2014
Subjunctive use variation among monolingual native speakers of Spanish: A cross-dialect analysis	Gallego y Alonso-Marks	Artículo de revista académ...	2014
On the Sequential Negotiation of Identity in Spanish-Language Discourse: Mobilizing Linguistic Resources in the...	Raymond	Tesis	2014
Frequency of Subjunctive Use in Oral Production in Two Varieties of Spanish	Gallego y Alonso-Marks	Artículo de revista académ...	2015
El subjuntivo español / su equivalencia en Checo	Fernán Bernán y Vales	Libro	2015
The Effectiveness of Collaborative Tasks on the Acquisition of the Spanish Subjunctive and Development of Mood...	Gallego	Artículo de revista académ...	2016
La selección modal del subjuntivo/indicativo en el Corpus sociolingüístico Prezee-Medellin	Koenig	Artículo de revista académ...	2016
A complete guide to the Spanish Subjunctive	Busch	Libro	2017
Heritage speakers' production and comprehension of lexically- and contextually-selected subjunctive mood mood...	Giancasprio	Tesis	2017
Acquisition of Spanish Mood Selection in Complement Clauses	Requena et al.	Sección de un libro	2017
Sociolingüística y pragmática del español	Silva-Corvalán y Enrique...	Libro	2017
Naciones cognitivas: clave para la descripción del nominal	Gómez Gavilho	Artículo de revista académ...	2018
Variation and grammaticalization in Romance: A cross-linguistic study of the subjunctive	Poplack et al.	Artículo de revista académ...	2018
Mood Choice and Context Availability: A Variationist Approach to the Subjunctive in New York City Spanish	Birnbaum	Tesis	2019
L2 Spanish morphosyntactic development through collaborative writing: An analysis of mood production, text le...	Gallego	Artículo de revista académ...	2019
Epileptic commitment and mood alternation: A semantic-pragmatic analysis of Spanish future-framed adverbials	Hoff	Artículo de revista académ...	2019
Selfhood and Mood Alternation: A Semantic-Pragmatic Analysis of Spanish Future-Framed Adverbials	Hoff	Tesis	2019
La adquisición de la expresión modal en español de L2	Galarza Galarza	Tesis	2020
Prácticas y concepciones sobre la alternancia verbal escrita en pasado irreal al comentar un texto literario en bach...	Hernández Rodríguez	Artículo de revista académ...	2020

The right-hand pane shows the details of the selected article, including the author 'Ernesto Hernández Rodríguez', the journal 'Revista Mexicana de Lingüística Aplicada a la Enseñanza de Lenguas', and the publication date '2020/04/16'.

FIGURA 3. ESTRUCTURA JERÁRQUICA DE LA INFORMACIÓN CONCENTRADA EN ZOTERO

Para guardar las referencias de los artículos citantes, los bibliógrafos pueden, en el caso ideal, valerse del complemento de Zotero para su navegador

de internet, pero, a veces, es necesario ingresar la referencia manualmente. En cualquier caso, siempre es indispensable verificar que los datos de la referencia agregada estén completos y sean correctos. Se pone especial atención en el tipo de documento citante, el orden correcto del nombre y apellidos del autor, el título, año de publicación y editorial.

Conforme se avanza en la lista de los documentos citantes, el bibliógrafo actualiza la cuenta de citas validadas en la hoja colaborativa o, en su caso, en las otras columnas de control («autocita», «sin acceso», etc.).

Respecto al proceso de validación de documentos citantes a más de un artículo analizado, y de citas hechas a artículos en coautoría, cabe hacer las siguientes precisiones:

Si un documento cita a dos o más artículos de la lista, este documento debe ser duplicado e integrado a la carpeta de Zotero de cada autor/artículo citado. Es importante recalcar que en Zotero debe existir un registro del documento citante en cada carpeta, y no un solo registro asociado a dos o más carpetas. Ejemplo: el artículo Y de Suárez Méndez cita a Martín Butragueño (2012) y a Barriga Villanueva (2010). El artículo Y de Suárez Méndez se integra tanto a la carpeta del artículo de Martín Butragueño (2012) como a la de Barriga Villanueva (2010).

Si un documento cita un artículo de la lista en co-autoría, este documento debe ser integrado a la carpeta de Zotero de cada autor. Ejemplo: Martín Butragueño y Barriga Villanueva escriben un artículo X en 2007. Un artículo Y cita dicho texto. El artículo Y se integrará tanto a la carpeta del artículo X en la colección de Martín Butragueño, como en la de Barriga Villanueva. En la hoja de cálculo colaborativa, el artículo X se encuentra en dos renglones: uno corresponde a la autoría de Martín Butragueño y otro a la de Barriga Villanueva, y en ambas filas se contabiliza la cita hecha por el artículo Y.

Si un documento Y del autor A cita al artículo X de ese mismo autor, cuenta como autocita para el autor A, pero como cita validada para sus coautores, de ser el caso. Ejemplo: Martín Butragueño y Barriga Villanueva escriben un artículo X en 2007. En 2008 Martín Butragueño, en el artículo Y, cita al artículo

X; así entonces, la cita en el artículo Y cuenta como cita para Barriga Villanueva, pero como autocita para Martín Butragueño.

Integración de la información y análisis. En esta sección se describen las herramientas utilizadas para integrar los insumos y hacer el análisis sistemático de los datos. Una vez que se ha terminado la revisión de las citas reportadas por GA, la BHD crea una carpeta para el año de análisis y, en ésta, crea un proyecto de RStudio (RSTUDIO TEAM, s/f). Además, genera una búsqueda de todas las citas por cada centro en Zotero. Los bibliógrafos verifican que ese número coincida con la suma de la columna *Citas Zotero XXXX* de la pestaña correspondiente a su centro en la hoja de cálculo colaborativa. Éste es uno de los procesos más demandantes, pues requiere de una revisión exhaustiva de cada carpeta en caso de que el número de citas guardadas en Zotero no coincida con la suma de citas en la hoja de cálculo. La BHD también revisa que todos los artículos citantes tengan: año de publicación (éste no puede ser mayor al año de análisis) y tipo de documento (no puede contener tipos de documentos no válidos). A continuación, exporta la carpeta Zotero de artículos citantes de cada centro en formato CSV a la carpeta donde se guarda el proyecto de RStudio. Usando un script de R, conjunta los CSV en uno solo, al que se añade la columna *Centro*, para saber a qué centro pertenece el artículo citado.

Con la hoja de cálculo colaborativa, la BHD agrupa a todos los centros en una misma pestaña, recuperando las siguientes columnas: *Autor, Título del artículo, Nombre de la revista, Editorial, Lugar de publicación, Año, Centro, Citas GA XXXX, Diferencias, Citas Zotero XXXX, Sin acceso XXXX, Autocita XXXX, No arbitrado XXXX, Otro XXXX y Liga GA*. Este archivo es analizado en OpenRefine (HUYNH & MAZZOCCHI, 2021) para asegurar que haya consistencia en los datos. Se uniforman las variables *Año* y *Lugar de publicación*, utilizando los distintos métodos de *clustering* disponibles en esta herramienta y se editan manualmente las inconsistencias no identificadas con estos métodos. La hoja de cálculo así revisada, que lista todos los artículos publicados y el número de citas validadas, se guarda en la carpeta del proyecto de RStudio.

Con los archivos de datos producidos a partir de la biblioteca colaborativa en Zotero y la hoja de cálculo en Google Drive, y usando scripts de R, la BHD lleva a cabo los análisis necesarios para generar los datos que se reportan a la Presidencia del COLMEX.

Consideraciones respecto al método

El uso de GA para identificar citas implica un proceso de minuciosa revisión y validación de datos, que se afina año con año. En nuestra experiencia con este método, hemos detectado algunas complicaciones generales que requieren la debida consideración. Se presentan a continuación siete advertencias que el equipo de la BDCV ha distinguido a la luz del análisis de 1.236 artículos publicados entre 2012 y 2020, las 6.561 citas que fueron identificadas por GA, y el proceso que se requiere para afinar los resultados derivados de esta información.

Inconsistencias y verificación. La primera advertencia es que no debe subestimarse la cantidad de tiempo que se debe invertir en verificar y rectificar las inconsistencias. Una de ellas está asociada a los nombres de autores hispanoamericanos. El algoritmo de GA parece lo suficientemente sofisticado para agrupar las variaciones de los nombres de un determinado autor y recuperar citas a pesar de las variantes con que se consignan los nombres en los documentos citantes. En contraste, el complemento del navegador para importar citas a Zotero frecuentemente importa los nombres hispanoamericanos alterando el orden de los elementos, y consigna el apellido materno como apellido paterno, y éste como un segundo o tercer nombre de pila. El uso de guiones para unir los apellidos maternos y paternos en algunas publicaciones mitiga este problema, pero es una práctica muy poco difundida aún. A menudo se requiere un trabajo de verificación y edición de los nombres de autor en Zotero.

Otra inconsistencia tiene que ver con los títulos repetidos o paralelos. Un autor citado puede presentar dos trabajos con un mismo título, o con títulos casi idénticos, que a menudo corresponden a distintas etapas de una misma

investigación, o retoman algún trabajo previo, como una tesis doctoral. En estos casos, los bibliógrafos pueden invertir tiempo considerable para identificar el trabajo citado. También ocurre que GA reporta como citante a un mismo artículo tantas veces como títulos paralelos encuentre. El ejemplo típico es el de los artículos indizados en SciELO, que a menudo tienen un título en dos o tres idiomas. En consecuencia, GA reporta un resultado por cada título encontrado en SciELO, pero únicamente se debe contar una cita, y consignar los otros resultados como repeticiones en la columna de control *Otro XXXX*. Este tipo de artículos demandan especial cuidado, y su monitoreo puede ser de los más extenuantes, y de los más proclives a causar errores de validación.

Falta de acceso. La segunda advertencia está relacionada con la posible falta de acceso al documento citante. Esto podría impedir la validación de muchas citas reportadas por GA. Es posible verificar las citas en documentos de acceso abierto y, en nuestro caso, muchos recursos pueden ser verificados gracias a las suscripciones de la BDCV, o a que forman parte de su acervo impreso. Sin embargo, es importante considerar que, aunque pocas (ver *Citas «no válidas»* abajo), hay publicaciones a las que no tenemos acceso, y cuyas citas no pueden validarse. Por tanto, la cantidad de recursos a los que una biblioteca tenga acceso puede limitar su posibilidad de validar citas reportadas por GA.

Variaciones por año de publicación y por disciplina. La tercera advertencia tiene que ver con la distribución desigual de los datos por año. Como se puede apreciar en la FIGURA 4, la cantidad de artículos publicados por año/centro fluctúa de manera irregular. Si bien podría decirse que se mantiene un promedio de 17 artículos por año, el 2015 es el año con el promedio más bajo (13,8 artículos por centro) y el 2013 el más alto (20,3).

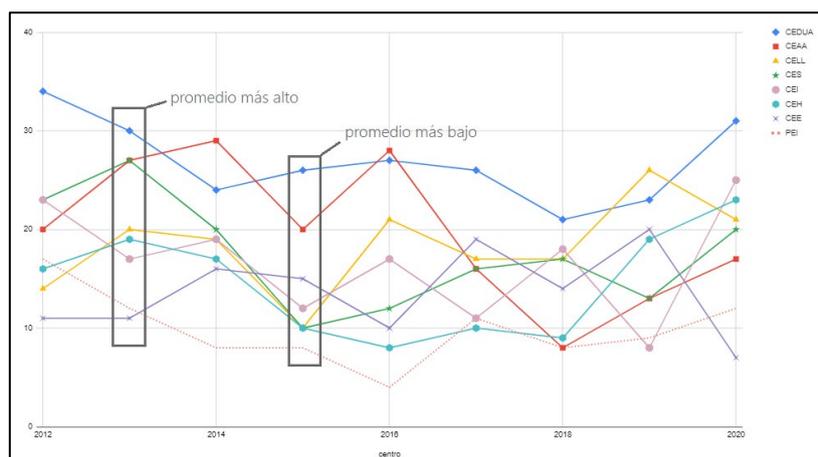


FIGURA 5. NÚMERO DE ARTÍCULOS POR CENTRO SEGÚN AÑO DE PUBLICACIÓN

Así, si un análisis como éste fuera a ser replicado en otra institución, recomendaríamos tomar una muestra de por lo menos cinco años para obtener las diferencias de comportamiento por año y por departamento. Con este último objetivo en mente, recomendamos visualizaciones como la FIGURA 5, cuyo diagrama de caja permite leer las siguientes características:

- La media por centro. Como puede observarse, ésta oscila entre 26 (CEDUA) y 9 (PEI) artículos por año.
- Las medidas más bajas y más altas por centro. Por ejemplo, para el CES, esta medida sería de aproximadamente 10 artículos por año, pues en 2015, 2016 y 2019 publicó entre 10 y 13 artículos por año; en contraste, en el caso del CELL, podemos ver un punto, que representa un valor atípico, pues sólo un año (en 2015) se publicaron 10 artículos.
- La dispersión de los datos. Vemos que la variación más alta es en el CEAA, cuya producción anual, en su mayoría puede ser descrita entre los 17 y 27 artículos anuales; mientras que el CELL tiene una producción más constante que oscila entre los 17 y 21 artículos por año.

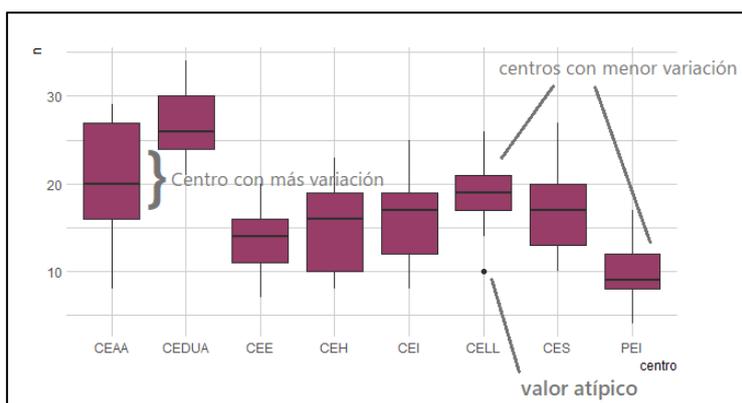


FIGURA 6. DIAGRAMAS DE CAJA DE NÚMEROS DE ARTÍCULOS POR CENTRO

Con datos como los de FIGURA 4, se pueden crear políticas institucionales que analizan las condiciones de los años más productivos, pero también que promuevan el establecimiento de una ventana de tiempo adecuada para el análisis de la producción. Ésta deberá establecerse después de verificar la variabilidad de los años con producción baja (para el caso de este estudio, el 2015) y alta (como el 2013). Por otro lado, la información que obtenemos con la FIGURA 5, nos permite tener un panorama de la variación por centro. Por ejemplo, mientras que los datos para un centro como el CELL o el PEI podrían ser significativos casi en cualquier rango de años (pues su variación no es tan extrema), no lo serían para centros como el CEAA o el CEH que tienen una variación más grande por año.

Citas «válidas». La cuarta advertencia corresponde a las complicaciones que existen para discernir e interpretar las citas válidas de las que detecta GA, en su dimensión temporal. En 2021, GA identificó 6.561 posibles citas a los artículos publicados entre 2012 y 2020. Después de la verificación manual 4.348 citas, es decir, un 66,27 % del total fueron consideradas válidas. Esta proporción varía desde un 53 % en 2020 hasta un 71 % en 2012. Es decir, nuestros datos sugieren que entre más tiempo haya pasado entre la publicación del artículo y el análisis de citas, mayor probabilidad existe de que la cita recuperada por GA sea una cita válida, pues el margen de error es menor a mayor cantidad de datos. Así, mientras el 47 % de citas no válidas en el 2020 equivale a unas cuantas citas (18 de un total de 39), el 71 % de citas válidas del 2012 equivale a 460 citas no válidas de un total de 1.562 (FIGURA 6).



FIGURA 7. CITAS SEGÚN GA POR AÑO DE PUBLICACIÓN DE ARTÍCULOS VS CITAS «VÁLIDAS»

En cuanto a la interpretación es importante considerar el correlato, es decir que la mayoría de las citas válidas corresponderá al extremo más antiguo de los datos. Por ejemplo, en nuestro análisis el 23 % de las citas identificadas por GA (es decir 1.562 citas) corresponden a las hechas a los artículos publicados en 2012, el primer año del periodo analizado; en contraste con el 2 % (39 citas) que es para los artículos de 2020 (ver FIGURA 6). Es decir, los artículos de 2012 han tenido ocho años para acumular citas; mientras que los de 2020, sólo tuvieron algunos meses o incluso días para ser descubiertos y citados por la comunidad académica.

Citas «no válidas». Como se ha mencionado, en el ejercicio de la BDCV además de filtrar los documentos no arbitrados, se han establecido otras categorías de citas «no válidas» cuyo desglose numérico es el siguiente: 1.094 duplicados, errores y otros (16 % del total de citas identificadas por GA y 46,8 % del total de no válidas); 609 se encontraban en documentos no arbitrados (9 % del total de citas identificadas de GA); 408 eran autocitas (6 % del total) y 226 citas provenían de documentos a los que no se tuvo acceso y no pudieron ser corroboradas (3 %).

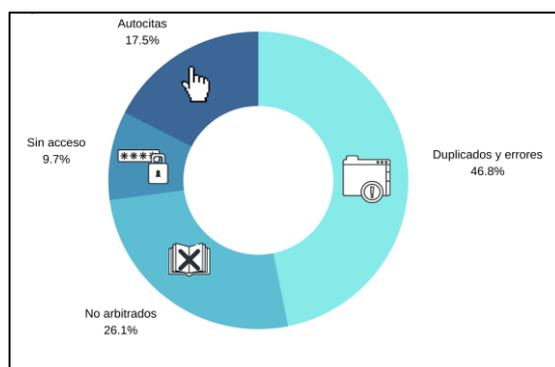


FIGURA 8. CITAS NO VÁLIDAS, PORCENTAJES POR TIPO

Esta es la primera vez, hasta donde sabemos, que se cuenta con una comprobación manual de los datos recuperados por el algoritmo de GA para artículos de CSyH y esperamos que contribuya a hacer lecturas más cautas sobre la información que este buscador maquinamente ofrece. Es decir, quien tome datos de GA para un análisis de citas a artículos de estas áreas mayoritariamente en español, puede calcular que habrá alrededor de un 16 % que corresponden a duplicados, errores y otros; cerca de un 9 % podría tratarse de documentos no arbitrados; y 6 % podrían ser autocitas. En consecuencia, si se aplican estos tres criterios, hasta un 32 % de las citas identificadas por GA serían citas no válidas. Más aún, como se puede leer en la FIGURA 6, esta proporción podría aumentar dependiendo del tiempo que han tenido para acumular citas. Así, la proporción de citas válidas en años más lejanos (ver 2012) podría ser mucho más alta que la de años más recientes en los que ha pasado poco tiempo para la acumulación de citas (ver 2020).

Documentos des-indexados. Si a las citas identificadas por GA que son 6.561, le restamos las citas válidas (4.348) obtenemos un total de 2.213 registros «no válidos». Sin embargo, si sumamos el total de «citas no válidas» de acuerdo con nuestro control, obtenemos 2.337 citas de este tipo. Es decir, existe una diferencia de 124 documentos «no válidos» que parecerían sobrar. Esto nos lleva a la sexta advertencia: hay documentos citantes que dejan de estar indexados en GA año tras año, lo cual se explica porque el índice de GA funciona automáticamente y si la cita no está escrita conforme a los

lineamientos que Google usa para su indización ese año, éstos no son incluidos (Google Scholar Help, s/f).

Por ejemplo, puede suceder que, en determinado año, el algoritmo de GA identifique que un artículo X tiene dos citas y que, al siguiente año, reporte cero, porque en el ajuste de su algoritmo dos artículos ya no son identificados como citantes. O bien, que en otro año se reporten dos citas y al año siguiente, cuatro, pero que estas cuatro sean todas nuevas, pues las del primer año ya no son identificados por el algoritmo como válidas. En el ejercicio que se hace en la BDCV, esto no es la norma, sino la excepción, pero calculamos que alrededor de 124 documentos citantes «no válidos» estaban presentes en búsquedas anteriores, pero han desaparecido en los resultados de GA.

Esto también se expresa en números negativos en la columna *Diferencias y Comprobación* de la hoja de control. Es decir, a veces al calcular la diferencia entre citas reconocidas por GA el año anterior al estudio para obtener las citas nuevas identificadas salen números negativos por lo explicado arriba. La práctica, en este caso, permite identificar que en GA hay registros que fueron validados en años anteriores y posteriormente ya no figuran entre los resultados. Esta volatilidad de los resultados de las citas que indiza GA, hace más compleja esta labor y justifica el uso de notas detalladas para el seguimiento de estos casos año tras año. Como si se tratará de una cadena de evidencia.

Diferencias por área. La séptima y última advertencia tiene que ver con los comportamientos de citación distintos de las Humanidades y las Ciencias Sociales. Si bien, de los 1.236 artículos publicados entre 2012 y 2020, 597 (el 48 %) fueron citados al menos una vez, este porcentaje varía entre los diferentes Centros.

Por ejemplo, los cuatro centros que pudieran considerarse más cercanos a lo que se denominan «Ciencias Sociales», son los que concentran la proporción más alta de citas a sus artículos publicados. El 73 % de los 123 artículos publicados por el profesorado del CEE fueron citados al menos una vez; está seguido por el CES, donde el 59 % de sus 158 artículos tuvieron por lo menos

una cita; en tercer lugar, está el CEDUA, que obtuvo una proporción ligeramente menor (57 % de sus 242 artículos) y por último, el CEI (46 % de sus 150 artículos). En contraste, los centros que pudieran considerarse más «humanistas», se encuentran en el extremo más bajo de la proporción: el CEH (45 % de 131), CELL (33 % de 165) y CEAA (28 % de 178).

Ahora, la advertencia va dirigida no sólo a reconocer esta diferencia sino a comprender que: primero, como se ha expresado en los reportes anuales que entrega la Biblioteca, la bibliografía apunta a que las Humanidades tienden a reportar pocas citas a artículos en revistas académicas, pues su circulación científica tiende a moverse más en materiales monográficos (LARIVIÈRE, GINGRAS, & ARCHAMBAULT, 2009; *Las revistas científicas y el español*, 2019), los cuales no han sido incluidos en este estudio. Segundo, la tendencia a publicar en línea también favorece la recuperación de citas hechas en artículos de revistas científicas; en contraste, un número (difícil de determinar) de citas hechas en libros son invisibles a una herramienta como GA, a menos que los motores de búsqueda de esta última tengan acceso a una versión legible por computadora del libro donde la posible cita aparece y esto afecta principalmente a las humanidades. Tercero, como se ha mencionado antes, un periodo tan corto (ocho años en nuestro estudio) afecta especialmente a las disciplinas de las humanidades, las cuales no suelen tratar temas de coyuntura y tardan en acumular citas.

Aún con todo, las proporciones obtenidas en nuestro estudio son alentadoras si se consideran las alarmantes cifras que DEREK CURTIS BOK recuperó en su libro sobre educación superior en Estados Unidos, en el que afirma que un 98 % de las publicaciones en humanidades y un 75 % de las de ciencias sociales, nunca obtienen una cita (BOK, 2013, p. 330). Este dato fue tomado de un artículo escrito 22 años antes (TAINER *et al.*, 1991) al que PENDLEBURY, en una carta abierta de ese mismo año, hizo una corrección numérica que no es menos preocupante: 93 % de los artículos de humanidades permanecen sin citas los primeros cinco años después de su publicación y entre 45 % y 49 % en el caso de las ciencias sociales (PENDLEBURY, 1991).

Sin embargo, considerando la misma ventana de tiempo (es decir, artículos que han tenido por lo menos cinco años para acumular citas), esa cifra jamás estuvo por arriba del 75 % de artículos sin cita para el caso de las Humanidades y el promedio de artículos con cita para ese periodo (2012-2015) fue de 34 % para el CEAA, 54 % para el CELL y 68 % para el CEH; es decir, siempre muy lejos de tener más del 93% de sus publicaciones sin cita como refiere Pendlebury (ver FIGURA 8).

Por otro lado, el de las Ciencias Sociales, aunque en 2014 el CEI tuvo su punto más bajo (47 % de artículos con citas) su promedio se mantuvo en 53 % de artículos con cita para ese periodo, CEDUA con 70 % de artículos con cita, y CEE con 83 %.

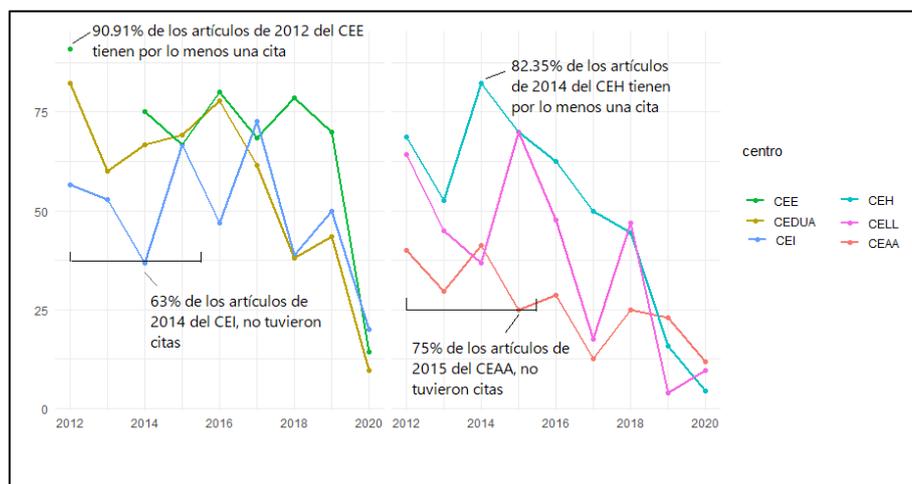


FIGURA 9. PORCENTAJE DE ARTÍCULOS CON POR LO MENOS UNA CITA

Conclusiones

En este artículo se ha descrito un método replicable para el análisis de citas utilizando GA como la mejor fuente abierta disponible para la identificación a artículos de CSyH en español. Hemos realizado una comparación ilustrativa de la dinámica de citación de la producción institucional de un tipo de documentos, los artículos. La dinámica en cuanto a los canales de difusión de cada disciplina es variable y obedece a lógicas de su propio campo de

producción. De tal modo que la medida justa para evaluar su desempeño debería tomar como referencia su propio campo.

Sin embargo, esto no quiere decir que esta sea la única vía y hemos mostrado las limitaciones de este método. Por un lado, quedado fuera del alcance de este estudio tratar iniciativas como [OpenCitations - Home](#) y [I4OC: Initiative for Open Citations](#) para aquellas instituciones que no tienen acceso a bases de datos de texto completo de paga para cotejar sus citas. Por otro lado, las limitaciones quedarán aún mejor definidas en cuanto se hagan estudios longitudinales para identificar cuantitativamente la diferencia de GA respecto de otras bases de datos, en relación con la cobertura, traslape y citas únicas para nuestro universo de artículos y en un futuro, la necesaria integración de otro tipo de publicaciones.

Cabe mencionar que este método es perfectible. Desde la BDCV ya existen esfuerzos para resarcir carencias, por ejemplo, crear una base de datos relacional en la que se guarden automáticamente los resultados de GA cada año para así disminuir, hasta cierto punto, el trabajo manual de comprobación, por lo menos, en el nivel de la comparación de documentos citantes con años anteriores y el registro de la cantidad de citas identificadas por los algoritmos de GA.

Por último, quisiéramos destacar que, a pesar de ser un procedimiento arduo, los análisis que se pueden derivar de este tipo de datos son sumamente ricos. Por ejemplo, en los informes anuales entregamos vistas de pájaro sobre la cobertura geográfica de los lugares de publicación de los diferentes centros, así como de las revistas que concentran la mayor cantidad de citas. Este tipo de información puede ser de utilidad para las instituciones académicas que desean tener una impresión de la cobertura geográfica de las citas hechas a los trabajos académicos, pero también para quienes deseen hacer análisis más finos de la circulación del conocimiento por disciplinas y encontrar áreas de oportunidad. En informes futuros, esperamos poder integrar análisis más robustos sobre la inmediatez de las citas, es decir, el tiempo que pasa para que un artículo reciba su primera cita.

Considerando lo anterior, advertimos que es necesario el uso crítico de fuentes para el análisis bibliométrico. Como hemos visto, el uso de una fuente limita o condiciona los resultados. Si bien no existe aún alguna fuente absolutamente confiable y exhaustiva para el análisis de citas a literatura en español y de CSyH, si seguimos utilizando las mismas fuentes como un criterio estático, sin analizar sus limitaciones, pero también sus posibilidades, mantendremos una visión aún más sesgada del rico universo de la comunicación científica que se da desde nuestras regiones y en nuestro idioma.

Bibliografía

- BOK, D. C. (2013). *Higher education in America*. Princeton University Press.
- GINGRAS, Y. (2016). *Bibliometrics and research evaluation: Uses and abuses*.
- GOOGLE SCHOLAR Help. (s/f).
<https://scholar.google.com/intl/en/scholar/inclusion.html#indexing>
- HARZING, A.-W., y ALAKANGAS, S. (2016). Google Scholar, Scopus and the Web of Science: A longitudinal and cross-disciplinary comparison. *Scientometrics*, 106(2), 787-804.
<https://doi.org/10.1007/s11192-015-1798-9>
- HUYNH, D., y MAZZOCCHI, S. (2021). OpenRefine (Versión v3.4.1) [Java].
<https://github.com/OpenRefine>
- LARIVIÈRE, V., GINGRAS, Y., y ARCHAMBAULT, É. (2009). The decline in the concentration of citations, 1900–2007. *Journal of the Association for Information Science and Technology*, 60(4), 858-862.
- Las revistas científicas y el español*. (2019). El Colegio Nacional. [video] <https://youtu.be/rd8-Rgr5bZE>
- MARTÍN MARTÍN, A., ORDUNA MALEA, E., THELWALL, M., y DELGADO LÓPEZ CÓZAR, E. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics*, 12(4), 1160-1177.
<https://doi.org/10.1016/j.joi.2018.09.002>
- PENDLEBURY, D. A. (1991). Science, citation, and funding. *Science*, 251(5000), 1410-1411.

- PLAZA, L. M., GRANADINO, B., GARCÍA CARPINTERO, E., ALBORNOZ, M., BARRERE, R., y MATAS, L. (2018). El valor del idioma español en ciencia y tecnología. *Rilce. Revista de Filología Hispánica*, 716-745. <https://doi.org/10.15581/008.34.2.716-45>
- ROVIRA, C., CODINA, L., y LOPEZOSA, C. (2021). Language Bias in the Google Scholar Ranking Algorithm. *Future Internet*, 13(2), 31. <https://doi.org/10.3390/fi13020031>
- RSTUDIO TEAM. (s/f). RStudio: Integrated Development Environment for R (Versión 1.4.1106). Boston, MA: RStudio, PBC. <http://www.rstudio.com/>
- STILLMAN, D. (2021). Zotero (Versión 5.0.96.2). Corporation for Digital Scholarship. <https://www.zotero.org/>
- TAINER, J. A., ABT, H. A., HARGENS, L. L., BOTT, D. M., LANCASTER, F. W., PANNELL, J. H., ... PENDLEBURY, D. A. (1991). Science, citation, and funding. *Science*, 251(5000), 1408-1411.

Eje 4: Infraestructura tecnológica

La importancia de los Repositorios Institucionales para el seguimiento automático de los Objetivos de Desarrollo Sostenible

JOSÉ DANIEL TEXIER

Universidad Nacional de Chilecito

jtexier@undec.edu.ar

JOHN ATKINSON

Universidad Adolfo Ibáñez

john.atkinson@uai.cl

RESUMEN

Los Objetivos de Desarrollo Sostenible (ODS) definidos en el 2015 por la Organización de las Naciones Unidas buscan lograr una sociedad mundial más equitativa. Sobre esta base, el objetivo de este trabajo es resaltar la importancia que tienen los Repositorios Institucionales (RI) para el seguimiento en la implementación de los ODS en un contexto deseado. Por tanto, se propone diseñar un clasificador computacional capaz de analizar los textos completos de las publicaciones científicas alojadas en los repositorios aplicando técnicas del Procesamiento de Lenguaje Natural para establecer vectores de similitud semántica con los ODS a partir de una metodología para la clasificación implementada con SciBERT y un *corpus* robusto seleccionado cuidadosamente. De esta manera, se le permite al usuario hacer uso de los recursos o ítems del repositorio o repositorios de acuerdo con la clasificación semántica lograda.

PALABRAS CLAVE

Procesamiento de lenguaje natural; ODS; RI; SciBERT; similitud semántica.

Motivación

En 2015, los 193 estados miembros de la Organización de las Naciones Unidas (ONU) adoptaron un plan para lograr un futuro mejor en la sociedad, trazando un camino durante los próximos quince años para poner fin a la pobreza extrema, proteger nuestro planeta y luchar contra la desigualdad y la injusticia (Naciones Unidas, 2015). En el corazón de la Agenda 2030 se encuentran los 17 Objetivos de Desarrollo Sostenible (ODS) que definen claramente el mundo que se desea, aplicándose a todas las naciones y sin dejar a nadie atrás. Además, el marco de indicadores globales incluye 169 metas y 232 indicadores distribuidos en las metas. Los nuevos Objetivos Globales son el resultado de un proceso que ha sido más inclusivo que nunca, con gobiernos involucrando a las empresas, la sociedad civil y los ciudadanos desde el principio. El cumplimiento de esta ambiciosa propuesta requerirá un esfuerzo sin precedentes por parte de todos los sectores de la sociedad, donde las empresas deben desempeñar un papel muy importante en el proceso. Los 17 ODS son (1) fin de la pobreza, (2) hambre cero, (3) salud y bienestar, (4) educación de calidad, (5) igualdad de género, (6) agua limpia y saneamiento, (7) energía asequible y no contaminante, (8) trabajo decente y crecimiento económico, (9) industria, innovación e infraestructura, (10) reducción de las desigualdades, (11) ciudades y comunidades sostenibles, (12) producción y consumo responsables, (13) acción por el clima, (14) vida submarina, (15) vida de ecosistemas y terrestres, (16) paz, justicia e instituciones sólidas, (17) alianzas para lograr los objetivos. Miles de millones de dólares de financiamiento están ahora vinculados a los ODS pero no está claro cómo se puede medir sistemáticamente el cumplimiento (GUPTA & VEGELIN, 2016). Para abordar este problema, la ONU publicó un nuevo conjunto de estándares para guiar a los administradores de fondos privados hacia inversiones que promuevan los ODS (GUPTA & VEGELIN, 2016). Esto tiene como objetivo acelerar significativamente la implementación para avanzar antes de la fecha límite de los objetivos en el 2030. La ONU proporciona guías prácticas que facilitan al sector privado la puesta en práctica de los ODS y acortan el camino desde el interés hasta la adopción de buenas prácticas de impacto (GUPTA & VEGELIN,

2016). En consecuencia, un mecanismo de revisión sólido para la implementación de la Agenda 2030 para los ODS requiere un marco sólido de indicadores y datos estadísticos para monitorear el progreso, informar las políticas y garantizar la rendición de cuentas de todas las partes interesadas.

Este plan de las Naciones Unidas surgió en la Era de la Información, lo que ha permitido replantear muchos procesos científicos gracias a la cantidad de datos e información que se genera cada segundo, minuto, hora y día, aunado a la generación de conocimiento. Evidencia de esto es el crecimiento de implementaciones de Repositorios Institucionales (Jisc, 2021), 4.948.726.800 de usuarios en internet para la fecha, tres mil millones de búsquedas, 600 millones de tweets, cinco mil millones de GB de datos, una totalización de 280 millones de registros en bases de datos bibliográficas elaborados por unos 35 millones de autores (principalmente, de artículos científicos y/o académicos) en 70 millones de revistas (CAMBIA.ORG, 2021; INTERNETLIVESTATS.COM, 2021; WEBSITEHOSTINGRATING.COM, 2020). Por tanto, este contexto obliga a plantearse cómo los Repositorios Institucionales (RI) pueden ayudar al seguimiento de los ODS en el mundo desde el punto de vista académico-científico sin entrar en los indicadores y datos estadísticos que exige la ONU para monitorear a los 193 estados miembros que adoptaron la agenda ODS en el 2015.

Al estudiar el seguimiento de los ODS desde los RI, la propuesta sienta sus bases en la relación de las Ciencias de la Computación y las Ciencias de la Información y Documentación, donde el desafío no está solo en la cantidad y calidad de los datos de los RI, sino en qué se debe hacer con todos ellos y cómo se gestionarán e integrarán con grandes volúmenes de información digital estructurada, semiestructurada y no estructurada, creados por personas y/u organismos en todos los ámbitos de la sociedad (GRIFFITHS *et al.*, 2016; SUBER, 2012). La forma de seguir el cumplimiento de los ODS por parte de la sociedad a través de los RI se hará implementando un clasificador a partir de los tres procesos que establece el concepto de la datificación para poder transformar tales datos de los RI en información susceptible de ser utilizada en cualquier área del conocimiento (LOPE SALVADOR *et al.*, 2020; MARKUS, 2017), es decir, el proceso de *recuperación* en obtener los artículos, el proceso de

transformación de los metadatos necesarios para ser analizados semánticamente de acuerdo con el Procesamiento de Lenguaje Natural (PLN) y el proceso de *visualización* para mostrar los resultados adaptadas a las necesidades del usuario para hacer el seguimiento de los ODS.

Entonces, para destacar la importancia de los RI en todo lo planteado, se propone desarrollar un modelo de clasificación con técnicas de PLN a partir de los recursos de los RI y de las 169 metas de los 17 ODS definidos por la ONU¹. El modelo será capaz de analizar el texto completo de los artículos científicos de los RI aplicando el PLN para establecer *matching* semánticos entre los recursos y los ODS de modo de establecer el cumplimiento y seguimiento por parte de la sociedad según lo publicado en los RI. Las contribuciones clave de este trabajo se resumen a continuación:

- Un relevamiento bibliográfico de la relación entre el PLN y los ODS.
- Una metodología para la clasificación de artículos científicos de acuerdo con los ODS en los RI.
- Una propuesta de un componente de software en RI para el seguimiento de los ODS de acuerdo con el PLN.
- La importancia o no de los RI para el seguimiento de los ODS.

La propuesta está organizada de la siguiente manera: la segunda sección describe los trabajos existentes para la clasificación automática de ODS, la tercera sección presenta una propuesta preliminar para la detección automática de ODS y la cuarta sección discute las conclusiones del avance.

Trabajos relacionados

Dada la cantidad de literatura disponible por las organizaciones que deben satisfacer los criterios de los ODS, es casi imposible analizarla manualmente con el fin de establecer si es que cumple uno o más criterios de los ODS. De ahí

¹ <https://sdgs.un.org/>

que el problema se debe abordar como una tarea que permita clasificar automáticamente grandes cantidades de textos temáticos en uno o más ODS. Algunos enfoques básicos utilizan análisis terminológicos de los documentos para detectar su relación con los ODS. Sin embargo, esto posee varios problemas relacionados con las sinonimias de términos, con la similitud semántica de diversos textos que corresponden a los mismos ODS y con la dificultad de caracterizar un documento y su relación con ODS solamente por los términos literales que contiene. Debido a esto, se requiere procesamiento de texto que no sólo abarque el nivel léxico y que considere repositorios de documentos completos de las diversas instituciones (BUTTIGIEG *et al.*, 2015; GALLI *et al.*, 2020; KOCH & KRELLENBERG, 2018; KÖRFGEN *et al.*, 2018, 2019; MESCHÉDE, 2020; MOLINA GÓMEZ *et al.*, 2020). A partir de una aclaración de las terminologías definidas por la ONU (BUTTIGIEG *et al.*, 2015), algunos enfoques (MOLINA GÓMEZ *et al.*, 2020) han utilizado técnicas de minería de textos a partir de un conjunto de palabras clave por cada ODS en artículos de prensa no estructurados (en internet). Un enfoque también basado en términos claves de los ODS se ha utilizado también en el área de la medicina (GALLI *et al.*, 2020). De igual manera, el proyecto UniNEtZ (KOCH & KRELLENBERG, 2018; KÖRFGEN *et al.*, 2018) realiza un análisis similar en textos pero en inglés y alemán sobre la base de una clasificación generada desde el proyecto. El aspecto en común de estos enfoques es que permiten generar términos que son las piezas fundamentales para la posterior construcción de una taxonomía ODS y disponer de un vocabulario controlado en la catalogación de los recursos existentes y por depositar al RI (TEXIER, 2013).

Debido a su simplicidad lingüística, los enfoques previos son incapaces de capturar información textual de más alto nivel sobre el contenido de las bases de datos que pueda relacionarse con los ODS. Sin embargo, algunos trabajos (JAYABALASINGHAM *et al.*, 2019) utilizan un enfoque de búsqueda por palabras clave de artículos científicos en bases de datos científicas como Scopus/Elsevier², donde dichas búsquedas las verifican y generan

² <http://scopus.com/>

actualizaciones en las palabras clave. De esta manera se incrementa la efectividad en la identificación de textos ODS. Otro aporte importante de este trabajo es el *matching* sobre una base de datos bibliográfica (i.e., Scopus) y que se tomó como base para la tercera edición *THE Impact Rankings*³. Una variación de este trabajo (BORDIGNON, 2021), aborda el problema de la polisemia de términos tanto como sea posible al enfocarse en las áreas temáticas más relevantes para cada ODS. Por otro lado, todos estos modelos tampoco procesan textos en RI o bibliotecas digitales ni consideran ese análisis bajo el punto de vista semántico (ALLEN *et al.*, 2018; BUTTIGIEG *et al.*, 2015; CASTRO *et al.*, 2020; GALLI *et al.*, 2020; KOCH & KRELLENBERG, 2018; KÖRFGEN *et al.*, 2018, 2019; MARTÍNEZ OSÉS, 2020; MESCHEDÉ, 2020; MOLINA GÓMEZ *et al.*, 2020).

El avance de la ciencia, como es natural, ha intentado abordar esta insuficiencia en la representación lingüística demostrada en los enfoques previos, utilizando modelos semánticos para la captura de información. Con este fin, algunos trabajos (HAJIKHANI & SUOMINEN, 2021; LAFLEUR, 2019; LUCCIONI *et al.*, 2020) introducen el uso de técnicas de aprendizaje automático y de Procesamiento del Lenguaje Natural (PLN) para determinar el grado de relevancia de los ODS en los textos deseados. Sin embargo, en trabajos posteriores (PUKELIS *et al.*, 2020; VÄLME & RENMARKER, 2021) se presentan propuestas semánticas que identifican los textos mejorando indudablemente la relación entre los textos y los ODS a partir de la aplicación del PLN a través de técnicas de aprendizaje automático tales como clasificadores Naïve Bayes, Máquinas de Soporte Vectorial (SVM), y modelos de lenguaje basados en Redes Neuronales profundas como BERT. Sin embargo, ninguno de estos trabajos de identificación semántica se aplican sobre el entorno de los repositorios institucionales ni bibliotecas digitales.

Debido a lo anterior, un pilar fundamental de la solución al problema de la identificación y seguimiento de ODS en repositorios institucionales, es el uso de mejores modelos de PLN, aquellos que vayan más allá del uso de términos, y que exploren estructura y características semánticas (GELBUKH & SIDOROV, 2010;

³ <https://www.timeshighereducation.com/rankings/impact/2021/overall>

JURAFSKY & MARTIN, 2008). Los enfoques modernos para abordar muchas tareas de PLN usualmente utilizan métodos probabilísticos o técnicas basadas en aprendizaje profundo (*Deep Learning*). Esto permite en muchas aplicaciones, entender y extraer información lingüística desde diferentes fuentes textuales, tales como RI, en diferentes niveles de conocimiento lingüístico (AGOGO & HESS, 2018; GELBUKH & SIDOROV, 2010; JURAFSKY & MARTIN, 2008; NARMADHA & SREEJA, 2016) a saber: léxico, sintáctico, semántico, pragmático, morfológico, discursivo.

Propuesta

Uno de los objetivos del trabajo es desarrollar un modelo de clasificación automática de los artículos de los RI que satisfacen criterios de los ODS. Para esto, el modelo debe poseer mayores capacidades de comprensión de lenguaje de modo que un método automático pueda aprender las mejores representaciones de los documentos, considerando la información estructural (por ejemplo: relación sintáctico-semántica entre términos de las oraciones de un texto), que finalmente permita clasificar dicha representación en algunas de las categorías de ODS predefinidas por la ONU.

Para lograr lo anterior, el enfoque propuesto consiste en los componentes que se observan en la FIGURA 1, en la cual los documentos científicos de un RI son analizados por un modelo de lenguaje basados en redes neuronales del tipo BERT que ha sido pre-entrenado para artículos científicos, denominado SciBERT (BELTAGY *et al.*, 2019). El resultado es un conjunto de vectores aprendidos (*word embeddings*) que representan las diferentes características (*features*) de los documentos. Posteriormente, estos alimentan un método de clasificación automática bayesiana multi-clase, el cual es capaz de determinar automáticamente las categorías más probables a las que pertenecen los documentos representados previamente.

Para lograr lo anterior, se deben realizar las siguientes tareas específicas:

- **Preparación del corpus de referencia:** se toma los documentos de los repositorios a evaluar, se preprocesan para cada uno de los 17 ODS y 169 metas. Los documentos que conforman el corpus de referencia son: 500 artículos por cada ODS que se obtendrán desde la base de datos de *Elsevier* (JAYABALASINGHAM *et al.*, 2019), Impact Rankings 2021 THE⁴, información oficial de la ONU⁵ y SDG Tracker⁶.
- **Entrenamiento del modelo de lenguaje basado en contexto (*word embeddings*):** permite generar los vectores de *features* que representan el contexto de cada uno de los documentos del repositorio de referencia. Para esto, se utilizará el modelo pre-entrenado del tipo BERT para literatura científica llamado SciBERT (BELTAGY *et al.*, 2019).
- **Clasificación de los documentos:** toma los vectores de *features* y entrena un clasificador multi-clase para determinar automáticamente el o los ODS que satisfacen cada uno de los documentos científicos de entrada que provienen del RI.

⁴ <https://www.timeshighereducation.com/world-university-rankings/impact-rankings-2021-methodology>

⁵ <https://www.un.org/sustainabledevelopment/poverty/>

⁶ <https://sdg-tracker.org/>

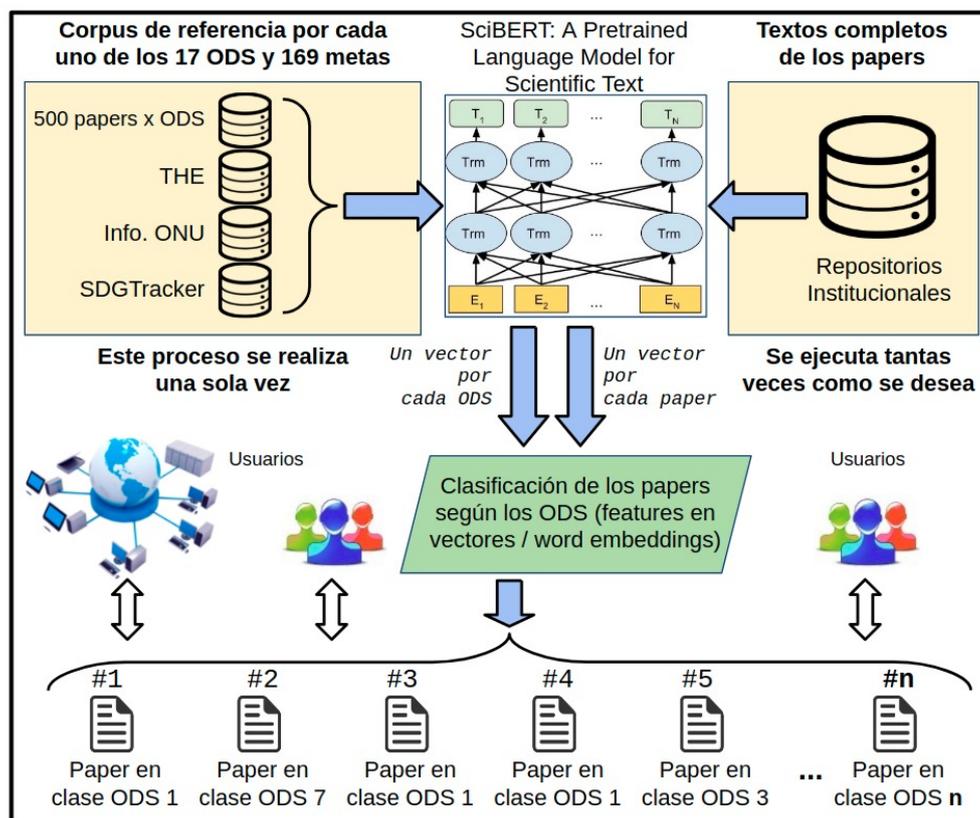


FIGURA 1. CLASIFICADOR DE CRITERIOS ODS

Elaboración propia

Conclusiones

Este trabajo presenta un enfoque para la detección automática de ODS en textos, que destaca la importancia de los RI para el seguimiento de los ODS. Para esto, se explora la relación entre los RI, la Era de la Información, los ODS y el Procesamiento de Lenguaje Natural. Según el directorio internacional de registros de repositorios OpenDOAR, existen 5.703 repositorios al 1º de julio de 2021 (Jisc, 2021), todos ellos son potenciales clientes para usar el clasificador y serán entes que van a permitir el seguimiento de los ODS a partir de sus trabajos científicos depositados en los repositorios.

A pesar de que existen varios trabajos para la detección de ODS en textos científicos, estos capturan información a un nivel básico por lo que son incapaces de identificar relaciones más complejas en el uso de las palabras, tales como la polisemia. A diferencia de estos, el trabajo propone un modelo que utiliza técnicas de PLN y un clasificador multi-clase que permiten no sólo

capturar información textual resolviendo tal tipo de problema lingüístico sino que reconociendo información estructural de un texto. Para esto, se utiliza un enfoque que aplica modelos de redes neuronales profundas pre-entrenados como SciBERT para literatura científica. Así el enfoque se compone de tres fases para detectar información relativa a ODS en repositorios institucionales: preparación de colecciones de textos de referencia, uso del modelo SciBERT para extraer vectores de características (i.e., *word embeddings*) y clasificación de documentos en ODS utilizando dichos vectores.

Finalmente, este trabajo es el primer paso para que cada RI que lo desee, pueda contar con un clasificador automático que servirá para etiquetar todos sus recursos, pero también servirá como una herramienta para identificar qué recursos se identifican con determinados ODS en los procesos de catalogación.

Bibliografía

- AGOGO, D., & HESS, T. J. (2018). Scale Development Using Twitter Data: Applying Contemporary Natural Language Processing Methods in IS Research. In A. V. DEOKAR, A. GUPTA, L. S. IYER, & M. C. JONES (Eds.), *Analytics and Data Science: Advances in Research and Pedagogy* (pp. 163-178). Springer International Publishing. https://doi.org/10.1007/978-3-319-58097-5_12
- ALLEN, C., METTERNICHT, G., & WIEDMANN, T. (2018). Initial progress in implementing the Sustainable Development Goals (SDGs): A review of evidence from countries. *Sustainability Science*, 13(5), 1453-1467. <https://doi.org/10.1007/s11625-018-0572-3>
- BELTAGY, I., LO, K., & COHAN, A. (2019). SciBERT: A Pretrained Language Model for Scientific Text. ArXiv:1903.10676 [Cs]. <http://arxiv.org/abs/1903.10676>
- BORDIGNON, F. (2021). Dataset of search queries to map scientific publications to the UN sustainable development goals. *Data in Brief*, 34, 106731. <https://doi.org/10.1016/j.dib.2021.106731>
- BUTTIGIEG, P. L., MCGLADE, J., & COPPENS, L. (2015). Clarifying terms in the SDGs: Representing the meaning behind the terminology. *2nd Meeting of the Inter-Agency Expert Group on Sustainable Development Goal Indicators*. <https://doi.org/10013/epic.155ebbec-cae2-4e84-8630-28a7feca5b0e>

- CAMBIA.ORG. (2021). *Cambia – Enabling Innovation*. <https://cambia.org/>
- CASTRO, F., VALBUENA, N., & CARIBE, C. DE LOS O. DE D. S. PARA A. L. Y EL. (2020). *Índice ODS 2019 para América Latina y el Caribe*. <https://bibliotecadigital.ccb.org.co/handle/11520/25484>
- GALLI, A., LEUENBERGER, A., DIETLER, D., FLETCHER, H. A., JUNGHANSS, T., & UTZINGER, J. (2020). Tropical Medicine and International Health and the 2030 Agenda for Sustainable Development. *Tropical Medicine & International Health*, 25(1), e1–e13. <https://doi.org/10.1111/tmi.13368>
- GELBUKH, A., & SIDOROV, G. (2010). *Procesamiento automático del español con enfoque en recursos léxicos grandes*. Instituto Politécnico Nacional.
- GRIFFITHS, M. D., KUSS, D. J., BILLIEUX, J., & PONTES, H. M. (2016). The evolution of Internet addiction: A global perspective. *Addictive Behaviors*, 53, 193-195. <https://doi.org/10.1016/j.addbeh.2015.11.001>
- GUPTA, J., & VEGELIN, C. (2016). Sustainable development goals and inclusive development. *International Environmental Agreements: Politics, Law and Economics*, 16(3), 433–448. <https://doi.org/10.1007/s10784-016-9323-z>
- HAJIKHANI, A., & SUOMINEN, A. (2021). The interrelation of sustainable development goals in publications and patents: A machine learning approach. *CEUR Workshop Proceedings*, 2871, 183-193. <https://cris.vtt.fi/en/publications/the-interrelation-of-sustainable-development-goals-in-publication>
- INTERNETLIVESTATS.COM. (2021). *Internet Live Stats—Internet Usage & Social Media Statistics*. <https://www.internetlivestats.com/>
- JAYABALASINGHAM, B., BOVERHOF, R., AGNEW, K., & KLEIN, L. (2019). *Identifying research supporting the United Nations Sustainable Development Goals*. 1. <https://doi.org/10.17632/87txkw7khs.1>
- JISC. (2021). *OpenDOAR: Directory of Open Access Repositories*. <https://v2.sherpa.ac.uk/opensoar/>
- JURAFSKY, D., & MARTIN, J. (2008). *Speech and Language Processing*. Prentice Hall.
- KOCH, F., & KRELLBERG, K. (2018). How to Contextualize SDG 11? Looking at Indicators for Sustainable Urban Development in Germany. *ISPRS International Journal of Geo-Information*, 7(12), 464. <https://doi.org/10.3390/ijgi7120464>
- KÖRFGEN, A., FÖRSTER, K., GLATZ, I., MAIER, S., BECSI, B., MEYER, A., KROMP-KOLB, H., & STÖTTER, J. (2018). It's a Hit! Mapping Austrian Research Contributions to the Sustainable Development Goals. *Sustainability*, 10(9), 3295. <https://doi.org/10.3390/su10093295>
- KÖRFGEN, A., GLATZ, I., MAIER, S., SCHERZ, M., KREINER, H., PASSER, A., ALLERBERGER, F., KROMP-KOLB, H., & STÖTTER, J. (2019). Austrian Universities and the Sustainable Development Goals. *IOP Conference Series: Earth and Environmental Science*, 323(1), 012156. <https://doi.org/10.1088/1755-1315/323/1/012156>
- LAFLEUR, M. (2019). *Art Is Long, Life Is Short: An SDG Classification System for DESA*

- Publications* (SSRN Scholarly Paper ID 3400135). Social Science Research Network. <https://doi.org/10.2139/ssrn.3400135>
- LOPE SALVADOR, V., MAMAQI, X., & BORDES, F. J. V. (2020). La Inteligencia Artificial: Desafíos teóricos, formativos y comunicativos de la datificación. *Icono14*, 18(1), 58-88.
- LUCCIONI, A., BAYLOR, E., & DUCHENE, N. (2020). Analyzing Sustainability Reports Using Natural Language Processing. ArXiv:2011.08073 [Cs]. <http://arxiv.org/abs/2011.08073>
- MARKUS, M. L. (2017). Datification, Organizational Strategy, and IS Research: What's the Score? *The Journal of Strategic Information Systems*, 26(3), 233-241. <https://doi.org/10.1016/j.jsis.2017.08.003>
- MARTÍNEZ OSÉS, P. J. (2020). Hacer realidad la Agenda: Medios de implementación, revisión y seguimiento. *Transformar Nuestro Mundo, ¿realidad o Ficción?*
- MESCHEDE, C. (2020). The Sustainable Development Goals in Scientific Literature: A Bibliometric Overview at the Meta-Level. *Sustainability*, 12(11), 4461. <https://doi.org/10.3390/su12114461>
- MOLINA GÓMEZ, N. I., RODRÍGUEZ, C. A., LÓPEZ, P. A., & DÍAZ ARÉVALO, J. L. (2020). *Minería de texto y aprendizaje automático para identificar prioridades de desarrollo sostenible*. <https://repository.usta.edu.co/handle/11634/23291>
- NACIONES UNIDAS. (2015). *Resolución de la Asamblea General de las Naciones Unidas: A/RES/70/1*. ONU. <https://undocs.org/es/A/RES/70/1>
- NARMADHA, R. P., & SREEJA, G. G. (2016). A survey on online tweet segmentation for linguistic features. *2016 International Conference on Computer Communication and Informatics (ICCCI)*, 1-6. <https://doi.org/10.1109/ICCCI.2016.7479955>
- PUKELIS, L., PUIG, N. B., SKRYNIK, M., & STANCIAUSKAS, V. (2020). OSDG – Open-Source Approach to Classify Text Data by UN Sustainable Development Goals (SDGs). ArXiv:2005.14569 [Cs]. <http://arxiv.org/abs/2005.14569>
- SUBER, P. (2012). Ensuring open access for publicly funded research. *BMJ: British Medical Journal*, 345. <https://doi.org/10.1136/bmj.e5184>
- TEXIER, J. (2013). *Los repositorios institucionales y las bibliotecas digitales: Una somera revisión bibliográfica y su relación en la educación superior*. 9. <http://eprints.rclis.org/19925/>
- UNITED NATIONS. (2015). *Resolution adopted by the General Assembly. A/RES/70/1*. United Nations Publications. <https://undocs.org/en/A/RES/70/1>
- VÄLME, E., & RENMARKER, L. (2021). *Accelerating Sustainability Report Assessment with Natural Language Processing*. <http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-445912>
- WEBSITEHOSTINGRATING.COM. (2020, January 4). *Best Web Hosting Companies In 2021 | Website Hosting Rating*. <https://www.websitehostingrating.com/>

Reportes estadísticos para repositorios digitales a partir de múltiples fuentes basado en el *stack* ELK

PABLO CÉSAR DE ALBUQUERQUE

PREBI-SEDICI, Universidad Nacional de La Plata

CESGI, Comisión de Investigaciones Científicas de la Provincia de Buenos Aires

pablo@sedici.unlp.edu.ar

GONZALO LUJÁN VILLARREAL

PREBI-SEDICI, Universidad Nacional de La Plata

CESGI, Comisión de Investigaciones Científicas de la Provincia de Buenos Aires

gonzalo@prebi.unlp.edu.ar

MARISA RAQUEL DE GIUSTI

PREBI-SEDICI, Universidad Nacional de La Plata

CESGI, Comisión de Investigaciones Científicas de la Provincia de Buenos Aires

marisa.degiusti@sedici.unlp.edu.ar

RESUMEN

Los Repositorios Institucionales (RI) tienen la necesidad de generar métricas de impacto que permitan comprender cómo son utilizados, lo que puede ayudar a generar información que permita asistir a la toma de decisiones; por ejemplo, fundamentar decisiones políticas o simplemente ofrecer servicios a los investigadores que depositan sus recursos y quieren saber cuál es la interacción del público con el repositorio. Actualmente, existen servicios que brindan la posibilidad de generar tableros de control a partir de los accesos registrados al repositorio a través del *tracking* de eventos y si bien estos servicios son muy utilizados tienen ciertas limitaciones al momento de cruzar información con otras fuentes de datos que pueden aportar información valiosa para analizar. En este trabajo se

utilizará el *stack* ELK para desarrollar un prototipo a partir de la base de datos del RI CIC-DIGITAL y se la asociará a los eventos registrados en el *log* de acceso al servidor donde se despliegan sus servicios, con el fin de generar tableros de control que ayuden a visualizar los recursos más accedidos en un intervalo de tiempo y poder interpretar cuáles son los factores que inciden en estos eventos.

PALABRAS CLAVE

Estadísticas; repositorio institucional; ELK. Statistics; Institutional Repository; ELK.

Introducción

Cada vez se necesitan más pruebas cuantitativas que ayuden a demostrar el valor de los servicios que hay en la web y la realidad que atraviesan los Repositorios Institucionales (RI) no difiere en este aspecto. Entre los tantos servicios que brinda un RI se encuentra el de la generación de métricas y estadísticas. Estas métricas pueden utilizarse para comprender mejor cómo se utilizan los repositorios, lo que puede ayudar a fundamentar las decisiones políticas sobre futuras inversiones y las decisiones de política técnica sobre las mejoras de la infraestructura técnica (KELLY *et al.*, 2012). También pueden ser de utilidad a la hora de tomar decisiones operativas a través de información que permita comprender los procesos involucrados en la difusión de los recursos almacenados en un repositorio, como por ejemplo:

- qué tipo de recursos se produce
- en qué áreas se investiga
- quiénes realizan estas investigaciones
- desde dónde
- en qué momento se generan los distintos recursos

- qué mecanismos se utilizan para producir o difundir los recursos
- cómo se utilizan tanto de manera interna como externa

Para medir muchos de estos compartimentos en los RI, se utilizan herramientas dedicadas al análisis web o Web Analytics (WA), entre los que mencionaremos a Matomo y Google Analytics debido a su amplia adopción.

Google Analytics

Google Analytics (GA) es un servicio gratuito utilizado por la mayoría de las bibliotecas académicas, basado en el etiquetado de páginas HTML para registrar la actividad de los visitantes en los servidores de Google. Su gran adopción se debe a que es muy fácil su integración, ya que sólo requiere registrarse en el sitio oficial, obtener un código de seguimiento (denominado Google Analytics Tracking Code) e insertarlo en cada una de las páginas que se desean analizar. Tiene la ventaja de que es muy personalizable no sólo desde el conjunto de paneles y alertas que ofrece sino que también permite definir dimensiones personalizadas. Sin embargo, algunos autores argumentan que GA es inapropiado para el uso educativo, ya que fue construido para el comercio electrónico y no para un entorno educativo (DRAGOŞ, 2011) y sumado a la pérdida de control de los datos debido a que son almacenados en servidores de terceros es que aparecen alternativas *open source* como Matomo.

Matomo

Matomo es una herramienta *open source* de análisis web que ofrece un servicio que permite evaluar todo el recorrido de los usuarios que visitan un sitio web. Muchas de sus funcionalidades están basadas en el servicio de analítica de Google, a punto tal que permite importar los datos sin perder el

histórico ya procesado. Para utilizarlo se requiere disponer de un servidor web, pero también cuenta con una versión *cloud* de pago.

El uso de las mencionadas herramientas para generar estadísticas es una buena alternativa para comenzar a generar métricas que permitan tener una visión más clara de lo que sucede en torno a un RI. Sin embargo, delegar esta responsabilidad no solo quita control sobre los datos que se manejan, sino fiabilidad en cómo se obtienen esas métricas. Por ejemplo en O'BRIEN *et al.*, (2016), los autores sostienen que hasta el 58 % de toda la actividad de RI generada por humanos no es reportada por Google Analytics. Otra de las limitaciones que presentan estas soluciones es la dificultad de integrar información proveniente de otras fuentes de datos como pueden ser la misma base de datos del repositorio, otros sistemas que formen parte la institución (como sistemas CRIS o portales de congresos o revistas), así como información generada por el uso del mismo sistema: accesos de usuarios, *logs* del servidor o incluso reportes de seguridad vinculados a cada sistema.

Propuesta

A partir de la problemática planteada, se realizó un prototipo a partir del stack de servicios open source llamado ELK, desarrollado por la empresa Elastic. Este stack de servicios se conforma de 3 herramientas que permiten integrar la información de diversas fuentes de datos en a partir de la implementación de pipelines, para luego realizar tableros de control a partir de los datos procesados.

Fuentes de datos

En este trabajo se utilizarán dos fuentes de datos para realizar los tableros de control. La primera fuente consiste en la base de datos del repositorio CIC-DIGITAL, mientras que la otra se trata del *log* de accesos al servidor sobre la cual se despliegan sus servicios.

El repositorio tomado para realizar estas pruebas está desarrollado sobre DSpace, un proyecto *open source*, que modela los recursos de la institución en

ítems y los organiza en una o más comunidades de nivel base que se organizan jerárquicamente en subcomunidades, como se puede observar en la FIGURA 1.

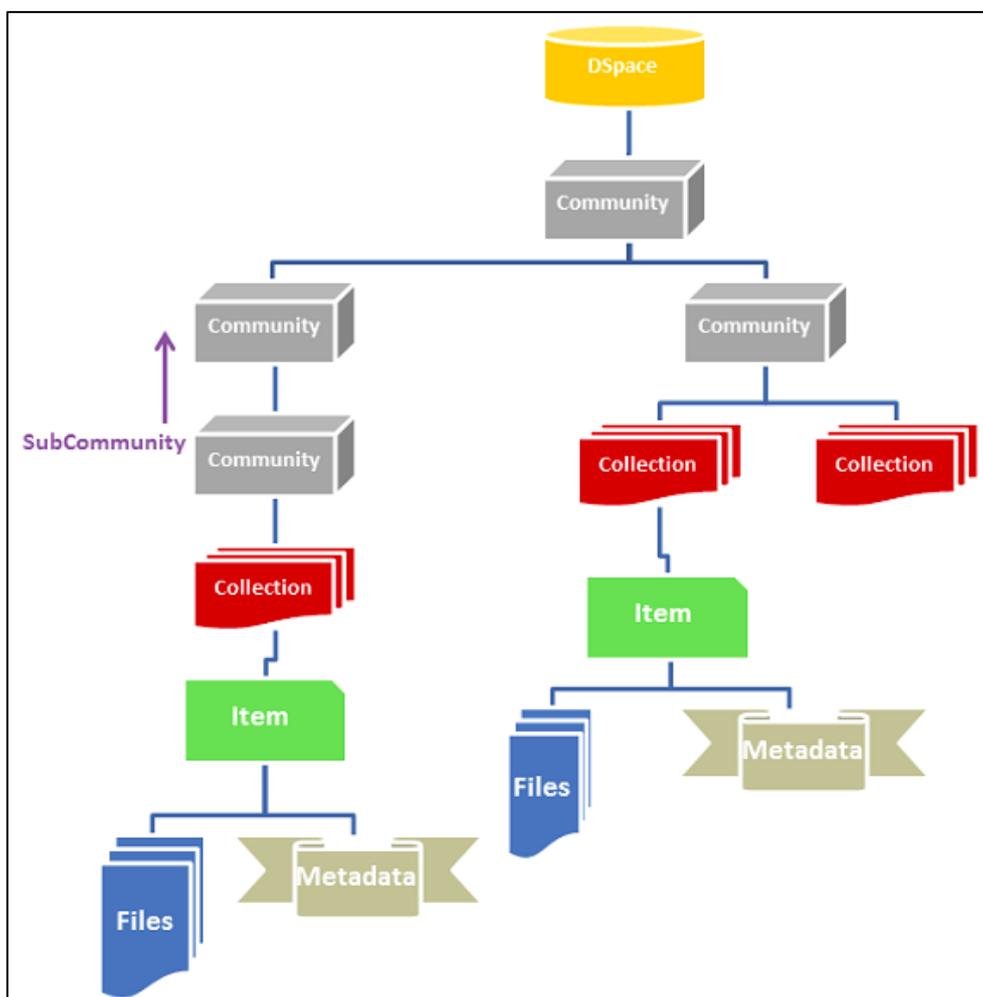


FIGURA 1

Fuente: *Functional Overview DSpace 6.x Documentation - LYRASIS Wiki (2018)*

En esta fuente de datos cada recurso, comunidad y colección del repositorio se asocia a una tabla de metadatos, que describen no solo los recursos que representan sino que también almacenan información de eventos, como la fecha en la que fue modificado un ítem por última vez, la fecha en la que ingresó al sistema y la fecha en la que el recurso se publicó. Otro aspecto a destacar es que tanto los ítems, como las comunidades y colecciones tienen asociado un identificador persistente, implementado en Handle, que será utilizado para poder interoperar con otras fuentes de datos. En la FIGURA 2 se

puede observar un diagrama del modelo de datos usado en DSpace, donde se puede apreciar la presencia del handle tanto en los ítems como en las comunidades y colecciones.

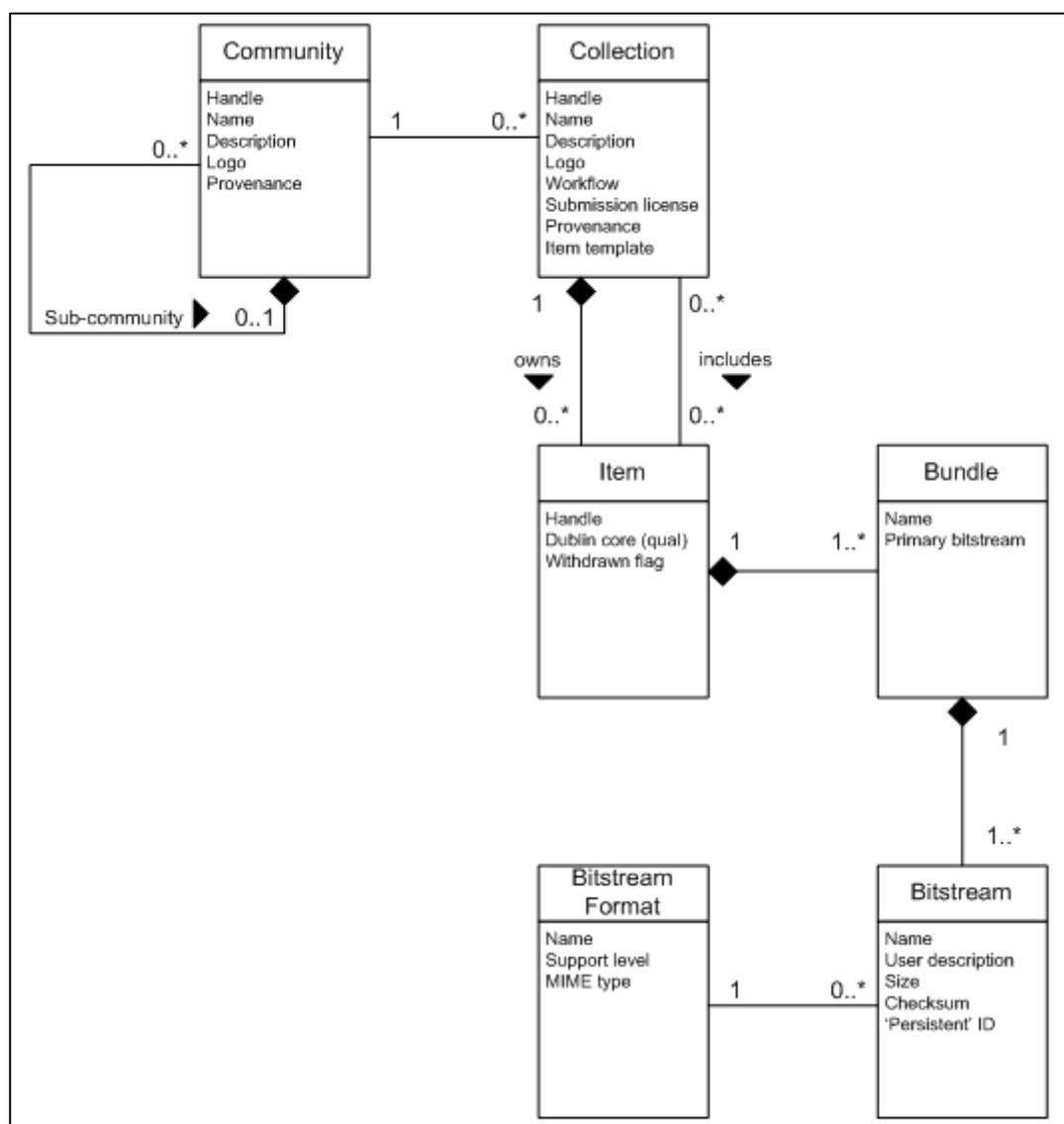


FIGURA 2

Fuente: *Functional Overview - DSpace 6.x Documentation - LYRISIS Wiki (2018)*

La base de datos utilizada en este trabajo está implementada sobre PostgreSQL, y en ella se modelan, entre otras cosas, los recursos, las colecciones, las comunidades y los metadatos que describen a los recursos.

Los servicios que ofrece este repositorio están desplegados en un servidor Apache, que registra sus accesos un archivo de *log* que será utilizado como segunda fuente de datos. En este archivo de *log*, se registra la IP que realiza la consulta al servidor, el User Agent del cliente, el código HTTP de la respuesta del servidor, el método HTTP con el que se realizó la consulta, el día, la fecha y la hora en que se atendió la petición, el tamaño de la respuesta y el recurso solicitado. En la Figura 3 se deja un ejemplo de cómo se registra un acceso al servidor en el archivo de *log*.

```
186.0.176.252 - - [17/Jun/2021:00:00:11 -0300] "GET
/bitstream/handle/10915/72076/Documento_completo.pdf?isAllowed=y&sequenc
e=1 HTTP/1.1" 200 7470507 "http://sedici.unlp.edu.ar/handle/10915/72076"
"Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML,
like Gecko) Chrome/91.0.4472.101 Safari/537.36"
```

FIGURA 3. EJEMPLO DE REGISTRO EN LOG

Stack ELK

Para realizar la integración de estas fuentes de datos se utilizará el *stack* ELK, desarrollado por Elastic, que se compone de tres partes. En primera instancia se encuentra Logstash, que será el encargado de recuperar la información de las diversas fuentes de datos y de procesarlas a través de un conjunto de *pipelines* definidos con el fin de normalizar los datos que finalmente se integrarán más adelante. Luego se encuentra Elasticsearch, un motor de búsqueda en el que se almacenarán los datos procesados por Logstash. Y finalmente se encuentra Kibana, una herramienta de visualización de datos que consume los datos indexados en Elasticsearch para generar diversos tableros de control.

Para poder levantar estos servicios se realizó un *fork* de un repositorio GitHub (LAPENNA, 2014/2021), que consiste de un proyecto *docker-compose* donde se definen cada uno de los servicios involucrados en el *stack* ELK, al cual se le realizaron algunas modificaciones, como definir nuevos servicios para las fuentes de datos. El código del proyecto se encuentra en [GitHub](#).

Para poder integrar los datos en Elasticsearch, en primera instancia se crearon dos índices para almacenar, por un lado los datos provenientes de PostgreSQL y por otro los logs de accesos. Estos índices llevan como nombre *items* y *logs*. La creación y la gestión de los datos de cada uno de estos índices es responsabilidad de instancias de Logstash independientes entre sí, que tomarán los datos correspondientes, los procesarán y los almacenarán en un índice en particular.

El índice de *items* contiene los siguientes campos:

- *@timestamp*: fecha en la que se indexó el documento
- *availabledate*: fecha en la que el ítem se publica en el repositorio
- *collection_handle*: handle de la colección a la que pertenece el ítem
- *community_handle*: handle de la comunidad a la que pertenece la colección
- *handle*: handle del ítem
- *hasfulltext*: si el ítem tiene texto completo
- *item_title*: título del ítem
- *last_modified*: fecha de la última modificación que sufrió el ítem
- *owning_collection*: colección a la que pertenece el ítem
- *subtype*: subtipo de recurso del ítem
- *type*: tipo de recurso del ítem

Los campos en el Índice *logs* son:

- *@timestamp*: fecha en la que se indexó el documento
- *agent*: User Agent del cliente
- *bytes*: tamaño de la respuesta expresado en bytes
- *clientip*: IP que realiza la consulta al servidor
- *handle*: handle del recurso accedido

- *referrer*: campo Referrer de la petición HTTP
- *request*: recurso solicitado al servidor
- *response*: código HTTP de la respuesta del servidor
- *timestamp*: día, fecha y hora en que se atendió la petición

Para recuperar los datos, las instancias de Logstash debieron usar distintos *plugins*. En el caso del Logstash dedicado a recuperar los datos de PostgreSQL, se utilizó el JDBC input plugin, mientras que para recuperar los datos del *log* de Apache, se utilizó Filebeats, un cliente ligero que es utilizado para enviar archivos a Logstash o Elasticsearch. Algo importante a aclarar es que en el Logstash que procesa los eventos de acceso se define un filtro, que a través del filtro *grok*, se encarga de obtener el *handle* a partir del *request*. Esto es importante ya que es a partir de este dato que se asociará un acceso al servidor con un ítem.

Una vez realizada esta primera carga inicial, una nueva instancia de Logstash se encarga de recuperar los datos del índice de *logs*, y en el proceso de filtrado se enriquecen estos registros a partir de los datos en el índice *items*. Para poder recuperar los datos se usó el Elasticsearch input plugin, mientras que la etapa de filtrado se realiza a través del filtro de Elasticsearch, que se encarga de asociar a partir del *handle* un evento de acceso con un ítem. Una vez realizada la asociación, esta instancia de Logstash depositará los eventos de acceso enriquecidos en un nuevo índice llamado *accesed-items* como se puede observar en la FIGURA 4.

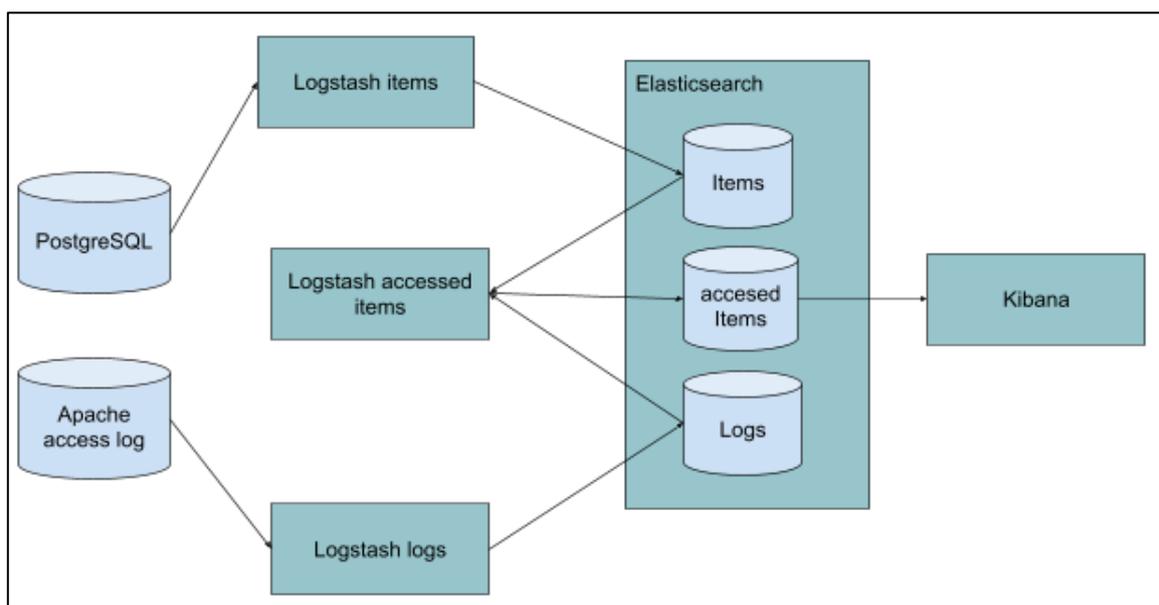


FIGURA 4

Una vez que se consolida el índice de *accessed-items* ya es posible realizar tableros en Kibana como el que se muestra a continuación.

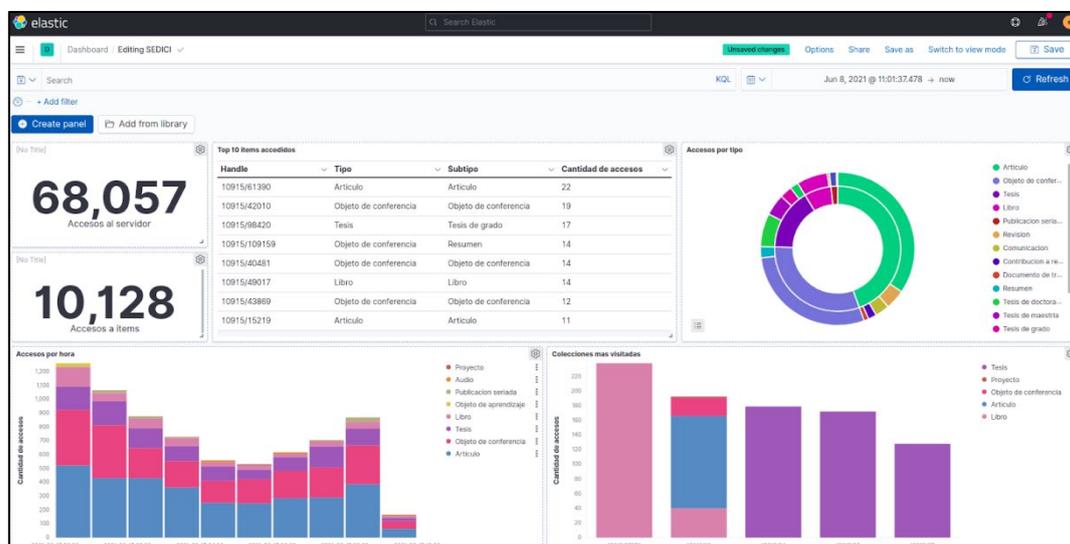


FIGURA 5. VISTA GENERAL DE TABLERO DE CONTROL

En la FIGURA 6 se puede observar dos métricas que muestran la cantidad de accesos que se han registrado en el archivo de *logs* y cuántos de esos accesos se corresponden con accesos a ítems, dejando de lado otras *requests*, como búsquedas en el repositorio o descargas de recursos.

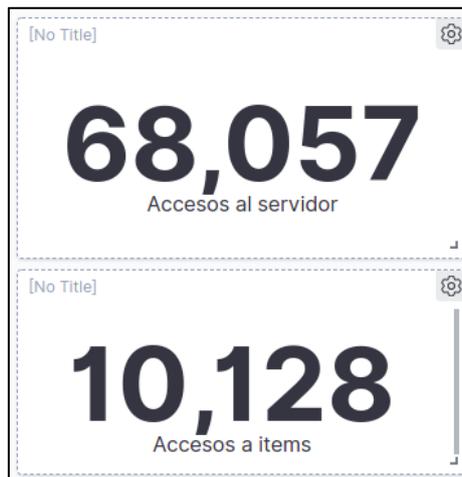


FIGURA 6. CANTIDAD DE ACCESOS AL SERVIDOR Y ACCESOS A ITEMS

Otra de las visualizaciones que se han generados es la de los 10 ítems más accedidos, junto con el tipo y subtipo de documento como se ven en la FIGURA 7.

Top 10 ítems accedidos			
Handle	Tipo	Subtipo	Cantidad de accesos
10915/61390	Articulo	Articulo	22
10915/42010	Objeto de conferencia	Objeto de conferencia	19
10915/98420	Tesis	Tesis de grado	17
10915/109159	Objeto de conferencia	Resumen	14
10915/40481	Objeto de conferencia	Objeto de conferencia	14
10915/49017	Libro	Libro	14
10915/43869	Objeto de conferencia	Objeto de conferencia	12
10915/15219	Articulo	Articulo	11

FIGURA 7. TOP TEN ITEMS ACCEDIDOS

También se realizaron dos visualizaciones para ver como es la distribución de tipos de documento. En la FIGURA 8 se puede observar un gráfico donde en el primer anillo se puede observar la cantidad de accesos por tipos de documentos, mientras que en el anillo externo se pueden visualizar los accesos por subtipos.

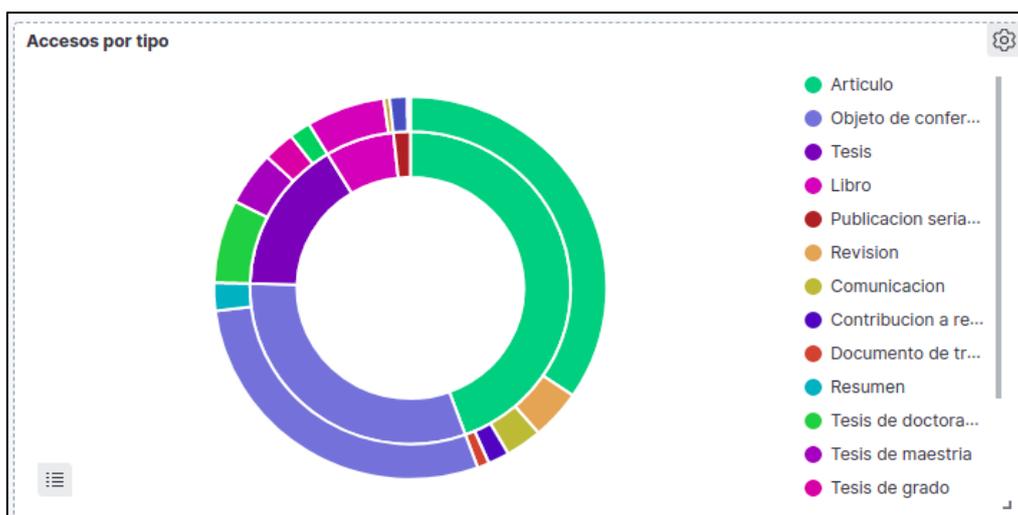


FIGURA 8. ACCESOS POR TIPO DE RECURSO

En la FIGURA 9 se pueden observar la cantidad de accesos por hora y la distribución de tipos de documentos.

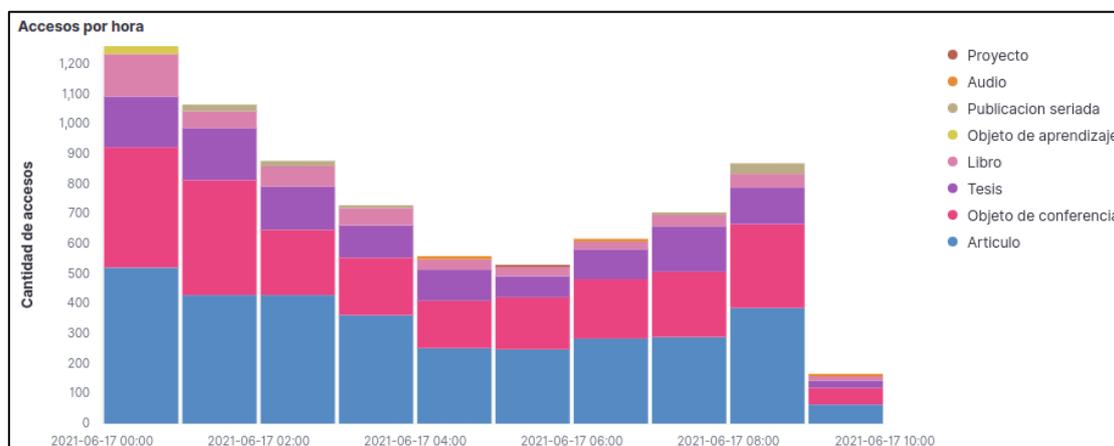


FIGURA 9. ACCESOS POR HORA

En la FIGURA 10 se muestra las colecciones más visitadas, a partir del handle de la colección.

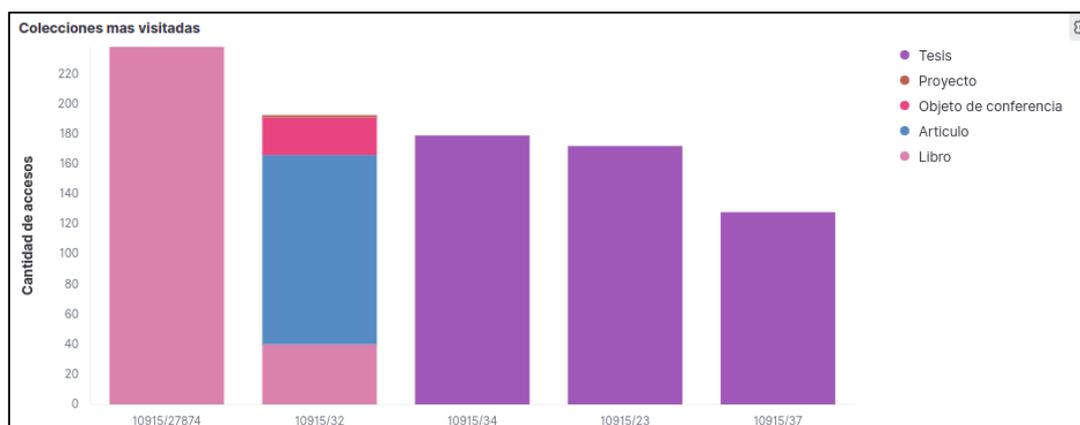


FIGURA 10. COLECCIONES MÁS VISITADAS

Conclusión

En este trabajo se han repasado algunas de las funciones y necesidades que atiende un repositorio, con énfasis en la generación de métricas y estadísticas destinadas a la asistencia en la toma de decisiones. Como se ha mencionado, el uso de herramientas como GA y Matomo, si bien provee muchas funcionalidades importantes para tener una visión más amplia de lo que sucede en un RI, la dificultad que se presenta al momento de integrar otras fuentes y a la falta de control sobre los datos (en el caso de GA) dan lugar a la implementación de soluciones como la propuesta, basada en el *stack* ELK.

En este caso, la información que se puede obtener a partir de estas visualizaciones puede ayudar a entender cómo la tipología de un recurso puede afectar en el interés de los usuarios de un repositorio, de forma tal que es posible realizar acciones como la de la promoción de los ítems más accesibles o bien ofrecer estas métricas como un nuevo servicio a los responsables del repositorio o autoridades de una institución.

Si bien los datos utilizados se corresponden a un lapso de tiempo muy acotado, sumado a que no muchos de los accesos registrados se corresponden con *bots* (como los utilizados por Google que indexan el sitio para ofrecer distintos servicios), resulta interesante cómo este *stack* permite definir diversos mecanismos de ingesta para cada una de las fuentes de datos

a utilizar, pudiendo generar métricas complementarias a las obtenidas por servicios de terceros, disponiendo siempre de los datos procesados.

Bibliografía

- DONOHUE, T. (2018). Functional Overview—DSpace 6.x Documentation—LYRISIS Wiki. <https://wiki.lyrasis.org/display/DSDOC6x/Functional+Overview>
- DRAGOŞ, S.-M. (2011). Why Google Analytics cannot be used for educational web content. 7th International Conference on Next Generation Web Services Practices, 113-118. <https://doi.org/10.1109/NWeSP.2011.6088162>
- KELLY, B., SHEPPARD, N., DELASALLE, J., DEWEY, M., STEPHENS, O., JOHNSON, G., & TAYLOR, S. (2012, julio 9). Open Metrics for Open Repositories.
- LAPENNA, A. (2021). Deviantony/docker-elk [Shell]. <https://github.com/deviantony/docker-elk> (Original work published 2014)
- OBRIEN, P., ARLITSCH, K., STERMAN, L., MIXTER, J., WHEELER, J., & BORDA, S. (2016). Undercounting File Downloads from Institutional Repositories. *Journal of Library Administration*, 56(7), 854-874. <https://doi.org/10.1080/01930826.2016.1216224>

Relevamiento de repositorios con perfiles de autor

EZEQUIEL MANZUR

PREBI-SEDICI, Universidad Nacional de La Plata

CESGI, Comisión de Investigaciones Científicas de la Provincia de Buenos Aires

ezemanzur@sedici.unlp.edu.ar

SANTIAGO TETTAMANTI

PREBI-SEDICI, Universidad Nacional de La Plata

CESGI, Comisión de Investigaciones Científicas de la Provincia de Buenos Aires

santit@sedici.unlp.edu.ar

GONZALO LUJÁN VILLARREAL

PREBI-SEDICI, Universidad Nacional de La Plata

CESGI, Comisión de Investigaciones Científicas de la Provincia de Buenos Aires

gonzalo@prebi.unlp.edu.ar

MARISA RAQUEL DE GIUSTI

PREBI-SEDICI, Universidad Nacional de La Plata

CESGI, Comisión de Investigaciones Científicas de la Provincia de Buenos Aires

marisa.degiusti@sedici.unlp.edu.ar

ARIEL JORGE LIRA

PREBI-SEDICI, Universidad Nacional de La Plata

CESGI, Comisión de Investigaciones Científicas de la Provincia de Buenos Aires

ariel.lira@sedici.unlp.edu.ar

RESUMEN

En este trabajo se realizó un relevamiento de los primeros 200 repositorios del Ranking Web de Transparencia en Repositorios Institucionales,

elaborado por Webometrics, con el fin de identificar aquellos que han implementado perfiles de autor para sus investigadores y destacar los servicios brindados. El relevamiento se dividió en dos etapas: la primera consistió en la búsqueda de aquellos repositorios dentro del ranking que brindan el servicio de perfiles de autor y por cada uno se relevaron varios datos como el nombre del repositorio, el tipo, la región y su URL; la segunda etapa consistió en, sobre el conjunto de perfiles de autor encontrados, un análisis de cada perfil implementado, destacando características como el tipo de servicios que ofrecen y cuál es el software que utiliza cada repositorio. Para realizar este análisis se destacaron siete puntos importantes a relevar en cada perfil: el software utilizado para el desarrollo del perfil, si dentro del perfil se muestran las publicaciones del autor, algún tipo de estadísticas, datos personales, datos sobre la filiación del autor, identificadores persistentes o algún enlace a las redes sociales académicas del autor. Además, se llevó a cabo una comparación entre los datos encontrados en este relevamiento y datos encontrados por análisis similares realizados en años anteriores por otros autores. Los resultados indican que un bajo porcentaje de los repositorios analizados provee el servicio de perfiles de autor, el 14 % (28) de los repositorios lo hacen. También que el software predominante entre los 28 repositorios que implementan dicho servicio es DSpace, y que los datos que mayormente se exponen en los perfiles son las publicaciones que realizaron los autores, información sobre la filiación y algunos datos personales, y luego, en menor medida, estadísticas, identificadores persistentes y las redes sociales académicas.

PALABRAS CLAVE

Perfiles de autor; repositorio institucional; identidad digital; Webometrics; redes sociales. Author profiles; Institutional Repository; digital identity; Webometrics.

Introducción

El perfil de autor es un conjunto de datos que recogen de forma estandarizada toda la actividad investigadora y la producción académica y científica de un autor (cargos desempeñados, proyectos de investigación, publicaciones realizadas, etc.) (ESCOBAR & PASTOR RUIZ, 2012). La implementación de este tipo de perfiles en un repositorio institucional trae numerosos beneficios a los autores que lo integran: ayudan a identificar unívocamente a un autor, mejoran su visibilidad y permiten mostrar su historial científico como investigador; es comparable a tener un CV abierto al mundo (GARCÍA PEÑALVO, 2017). A su vez, se considera que los perfiles son muy importantes para el propio repositorio, ya que promueven el autoarchivo, aumentan la visibilidad de la institución, facilitan el acceso a la producción y ofrecen una alternativa desde la misma institución a plataformas comerciales como ResearchGate o Academia.edu, al proveer un servicio de valor añadido que incentiva la participación y el uso del repositorio (GENOVÉS, 2017; BARRUECO CRUZ & NAVALÓN, 2015).

El relevamiento realizado en este trabajo se centró en páginas dentro de los repositorios dedicadas solamente a un autor que, o bien contribuyó con algún artículo o trabajo científico al repositorio, o bien forma parte de la institución en la que se aloja este. Estas páginas normalmente tiene información personal del investigador (como mail e incluso una breve biografía), sus trabajos publicados en el repositorio, las organizaciones a las que pertenece, alguna foto (o espacio para incluir una), identificadores persistentes y estadísticas de citas, descargas y visualizaciones de los trabajos del investigador (GENOVÉS, 2017).

Como antecedente, en febrero de 2017, de los primeros 50 repositorios institucionales mejor puntuados en el Ranking Web de Repositorios de Webometrics, sólo cinco de ellos (el 10 % de los repositorios del top 50) proveían el servicio de perfiles de autor. De esos cinco repositorios, tres estaban desarrollados con el software para repositorios DSpace y los dos restantes mediante Digital Commons (GENOVÉS, 2017).

Objetivos

El objetivo principal de este trabajo es obtener un panorama general de la implementación del servicio de perfil de autor en los repositorios institucionales, estimar el porcentaje de repositorios que implementan este servicio y evaluar si ha aumentado o disminuido a lo largo del tiempo. Asimismo se plantea indagar sobre cuáles son los servicios más utilizados en este tipo de perfiles, qué información se expone del autor y si realizan alguna integración con identificadores persistentes, y realizar un análisis cuantitativo del tipo de software utilizado.

Metodología

Es posible dividir la revisión llevada a cabo en dos etapas. En la primera de ellas se realizó el relevamiento de los primeros 200 repositorios del ranking web de transparencia de Webometrics, y a partir de dicho relevamiento se generó un reporte con aquellos repositorios que contaban con el servicio de perfil de autor. En la segunda etapa, para cada repositorio del reporte anterior, se realizó un análisis pormenorizado de las características del servicio de perfiles de autores ofrecido. Este análisis se dividió teniendo en cuenta los distintos servicios que suele ofrecer un perfil de autor:

- Publicaciones mostradas
- Estadísticas
- Redes sociales académicas
- Datos personales
- Datos de la filiación institucional del autor
- Software en el que se desarrolló el servicio
- Identificadores persistentes de autor

De cada uno de estos servicios se registró su presencia o ausencia en los perfiles de autores de los distintos repositorios relevados; en los casos en los que estaban presentes, se incluyó un mayor detalle acerca de las características principales del servicio y particularidades de la implementación propia de ese repositorio.

La elección del ranking web de transparencia de Webometrics, como base de datos para la investigación, se basó en que, si bien el ranking no brinda información acerca de la excelencia de las instituciones o la calidad de los repositorios que lo integran, sí expone los repositorios que poseen una mayor presencia y visibilidad web (AGUILLO *et al.*, 2010; *Webometrics Ranking Web of Repositories / Objectives*, s. f.); sumado a esto, el volumen de la información que integra este ranking es lo suficientemente amplia para poder incluir en el análisis repositorios de naturaleza muy diversa (universitarios, temáticos, de mayor o menor tamaño y pertenecientes a instituciones de distintos lugares del mundo).

Primera etapa: repositorios con perfiles de autor

De los 200 repositorios analizados, 28 (14 %) tienen implementado un servicio de perfil de autor para al menos uno de sus autores; mientras que en los primeros 50 repositorios de la lista se encontraron diez con esta característica, el doble de los que había en 2017 (GENOVÉS, 2017).

Este es el listado de esos 28 repositorios junto a su tipología, región geográfica y URL:

Repositorio	Tipo	Región	URL
Europe PubMed Central	Preprints / Agregador	Global/Europa	https://europepmc.org/
Research Papers in Economics	Preprints / Temático	Global	https://ideas.repec.org/
PhilPapers	Agregador / Temático	Canadá	https://philpapers.org
Utrecht University Repository	Institucional	Países Bajos	https://www.narcis.nl/
Università degli Studi di Milano Institutional Archive of Research	Institucional	Italia	https://air.unimi.it/
National Chiao Tung University Institutional Repository	Institucional	Taiwán	https://ir.nctu.edu.tw/
Digital CSIC	Institucional	España	https://digital.csic.es/
RiuNet Repositorio Institucional Universidad Politécnica de Valencia	Institucional	España	https://riunet.upv.es/
Seoul National University's Institutional Repository S-Space	Institucional	Corea del Sur	http://s-space.snu.ac.kr
Universitas Muhammadiyah Surakarta Digital Library	Institucional	Indonesia	http://library.ums.ac.id/
Universitat Autònoma de Barcelona Dipòsit Digital de Documents	Institucional	España	https://ddd.uab.cat/
Dépôt Institutionnel de l'Université Catholique de	Institucional	Bélgica	https://uclouvain.be/

Repositorio	Tipo	Región	URL
Louvain			
University of Ghent Institutional Archive	Institucional	Bélgica	https://biblio.ugent.be/
Xi'an Institute of Optics and Precision Mechanics CAS Institutional Repository	Institucional	China	http://opt.cas.cn/
Vietnam National University Digital Repository	Institucional	Vietnam	http://repository.vnu.edu.vn/
Kazan Federal University Repository	Institucional	Rusia	https://repository.kpfu.ru
Korea Advanced Institute of Science and Technology Open Access Self-Archiving System	Institucional	Corea del Sur	https://koasas.kaist.ac.kr/
Queensland University of Technology Institutional Repository	Institucional	Australia	https://eprints.qut.edu.au
Repositorio Institucional Universidad de Valladolid	Institucional	España	http://uvadoc.uva.es/
Utah State University Digital Commons	Institucional	Estados Unidos	https://digitalcommons.usu.edu/
Technical University of Denmark Online Research Database in Technology	Institucional	Dinamarca	https://orbit.dtu.dk/
CONICET Digital	Institucional	Argentina	https://ri.conicet.gov.ar/
University of Iowa Research Online	Institucional	Estados Unidos	https://ir.uiowa.edu/

Repositorio	Tipo	Región	URL
Archive Ouverte Université de Genève	Institucional	Suiza	https://archive-ouverte.unige.ch/
University of Massachusetts Amherst Scholarworks	Institucional	Estados Unidos	https://scholarworks.umass.edu/
Biblos-e Archivo Universidad Autónoma de Madrid	Institucional	España	https://repositorio.uam.es/
University of Pennsylvania ScholarlyCommons	Institucional	Estados Unidos	https://repository.upenn.edu/
Keio University Academic Resource Archive	Institucional	Japón	http://koara.lib.keio.ac.jp/

Segunda parte: análisis de los perfiles de autor

Software utilizado

Luego de relevar los repositorios que implementan perfiles de autor, se realizó un análisis del tipo de software utilizado, tanto para los repositorios digitales como para implementación de los perfiles de autor. De los 28 repositorios relevados, se destaca el uso del software DSpace (9) y de la plataforma Digital Commons (4); se identificaron también muchos sistemas con una única implementación, como DSpace Cris, Eprints, Drupal y desarrollos propios, entre otros (ver FIGURA 1).

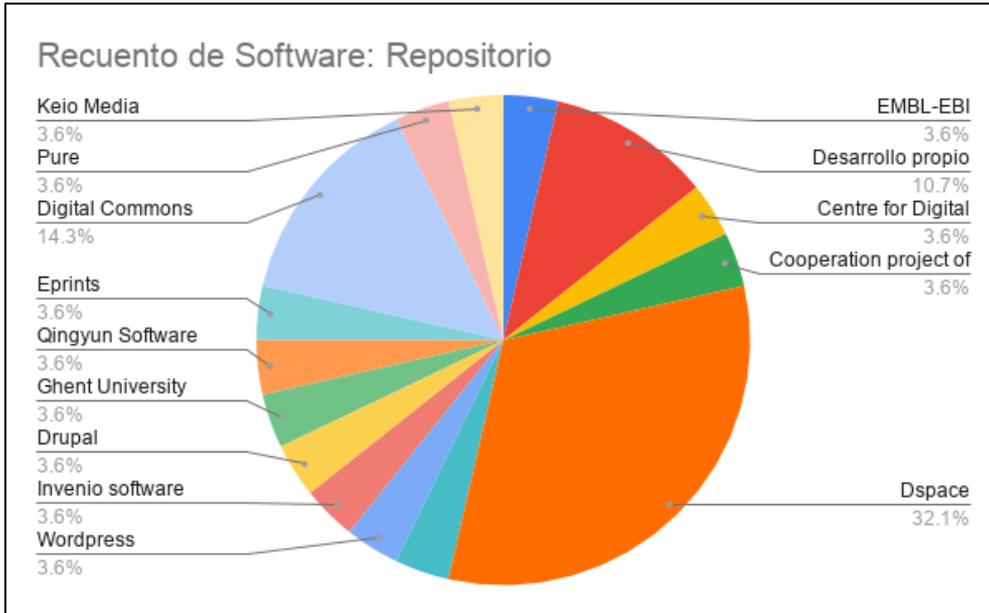


FIGURA 1. PORCENTAJE DE LOS REPOSITARIOS ANALIZADOS AGRUPADOS POR EL SOFTWARE EN EL QUE ESTÁN DESARROLLADOS

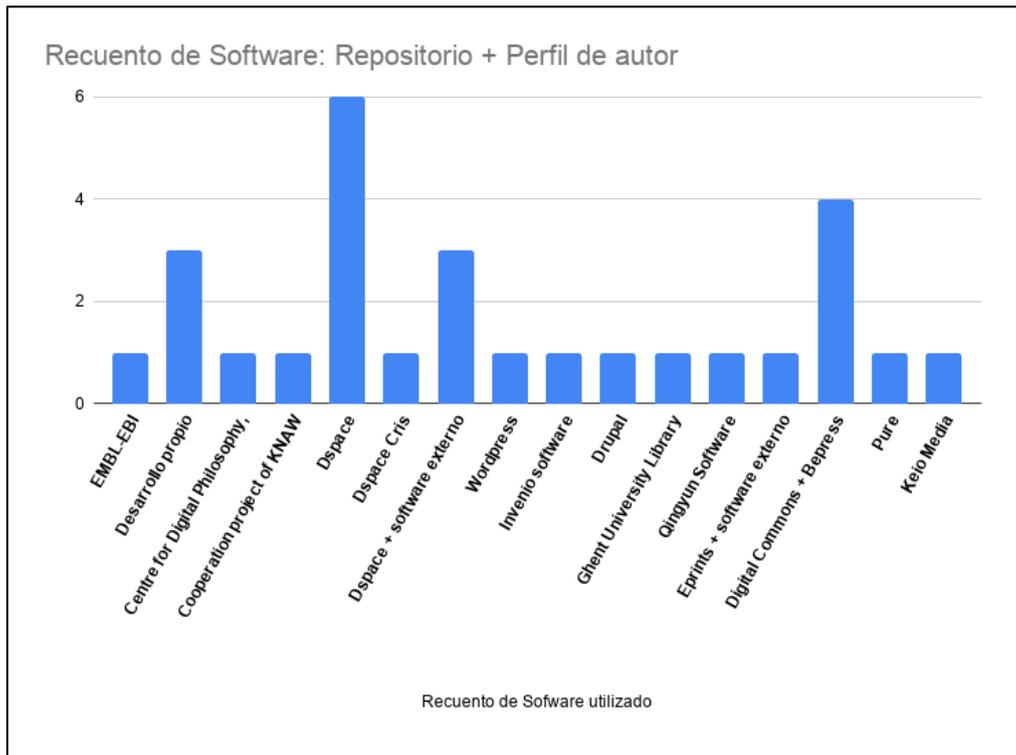


FIGURA 2. CANTIDAD DE REPOSITARIOS AGRUPADOS POR EL SOFTWARE EN EL QUE ESTÁN DESARROLLADOS JUNTO CON EL SOFTWARE CON EL QUE IMPLEMENTARON LOS PERFILES DE AUTOR

Publicaciones

La visualización de las publicaciones de un autor es una característica común a la mayoría de los repositorios. La totalidad de los 200 repositorios analizados permiten, en su servicio de búsqueda, filtrar las publicaciones por nombre de autor. Asimismo, en casi todos los perfiles de autor encontrados en el relevamiento (en 26 de los 28 repositorios con perfiles) se permite listar las publicaciones del investigador.

Sin embargo, si el listado de las publicaciones se realiza solo a partir del nombre del autor se podría incurrir en ambigüedades: varias personas pueden compartir el mismo nombre, por lo que el listado podría contener publicaciones de distintos autores (LATIF *et al.*, 2018). En comparación con el filtrado a partir de una búsqueda, el listado de publicaciones realizado a partir de un perfil de autor, reduce las chances de cometer estos errores, ya que normalmente se tiene mucha más información del investigador (filiación, áreas de investigación, co-autores), y los administradores de los repositorios pueden reconocer, con mayor facilidad, si el autor de una publicación es el mismo que el de un perfil dado, ya que además de su nombre, es posible incluir en el análisis información como las áreas de estudio y la filiación del autor, y comparar esta información con los datos de la publicación. Sumado a esto, en el caso de que un autor esté vinculado a un identificador persistente, es posible recopilar sus publicaciones a partir de dicho identificador y de este modo eliminar todo tipo de ambigüedad con su identidad (ESCOBAR & PASTOR RUIZ, 2012).

Del relevamiento surgió también que algunos perfiles de autor ofrecen servicios en conjunto con el listado de publicaciones del autor. Entre esos servicios caben destacar:

- Exportación de las publicaciones en distintos formatos (CSV, JSON, PDF); un ejemplo de esto ocurre en los perfiles de investigador del Repositorio de la Universidad de Utrecht.
- Filtros por co-autor, filiación y fecha, entre otros.
- Posibilidad de recomendar o «likear» la publicación de un autor, como en el caso del repositorio PhilPapers.

- Reporte de estadísticas sobre las publicaciones, de los que se hablará con detalle más adelante.
- Opción para exportar como bibliografía toda la obra del autor; este es el caso del repositorio de la Universidad de Ginebra.

Estadísticas

El uso de las visualizaciones de estadísticas es disímil a lo largo del listado, ya que once de los repositorios proveen este tipo de servicio, mientras que en los restantes diecisiete repositorios no se observaron estadísticas en los perfiles de autor. En aquellos casos en los que sí se incluyen estadísticas, la métrica más expuesta corresponde al número de visualizaciones y descargas de las publicaciones del autor; muchas veces estas métricas se exponen agrupadas por año o de acuerdo a la localización geográfica (ciudad, país o región). También se encontraron estadísticas del número de publicaciones del autor agrupadas por tipo, cantidad de publicaciones con los distintos co-autores y del número de citas al autor. Por lo general están acompañadas por distintos gráficos (algunos incluso interactivos), como se ejemplifica en la FIGURA 3, en donde se muestran gráficos de publicaciones agrupadas por tipo, cantidad de descargas y visualizaciones de sus publicaciones para un autor del repositorio de la Universidad Nacional de Vietnam; o mapas, en el caso de las estadísticas agrupadas por región geográfica, como se ve en los perfiles de autor del repositorio de la Universidad de Estudios de Milán (ver FIGURA 4).

Cabe destacar el caso del repositorio de la Universidad Autónoma de Madrid, el cual, de todos los repositorios relevados, es el que muestra una mayor cantidad y diversidad de estadísticas. Entre ellas podemos nombrar: índices biométricos en las publicaciones (índices h, q1, d1 y cantidad de citas), líneas temporales con la información sobre la formación académica y la trayectoria profesional, destaca los países en donde el autor presentó publicaciones en algún congreso y el tipo de presentación realizada (póster, ponencias, conferencias), impacto de sus publicaciones en redes sociales e incluso una red de colaboración institucional, donde se expone el rol del autor en los

proyectos de la universidad en los que participó, junto con el nombre y el rol de los distintos colaboradores (ver FIGURA 5).

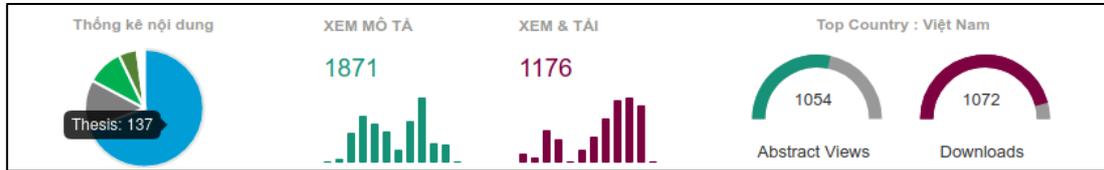


FIGURA 3. GRÁFICOS DE PUBLICACIONES AGRUPADAS POR TIPO, CANTIDAD DE DESCARGAS Y VISUALIZACIONES DE SUS PUBLICACIONES PARA EL AUTOR «NGUYỄN HUY CHƯƠNG» EN EL REPOSITORIO DE LA UNIVERSIDAD NACIONAL DE VIETNAM

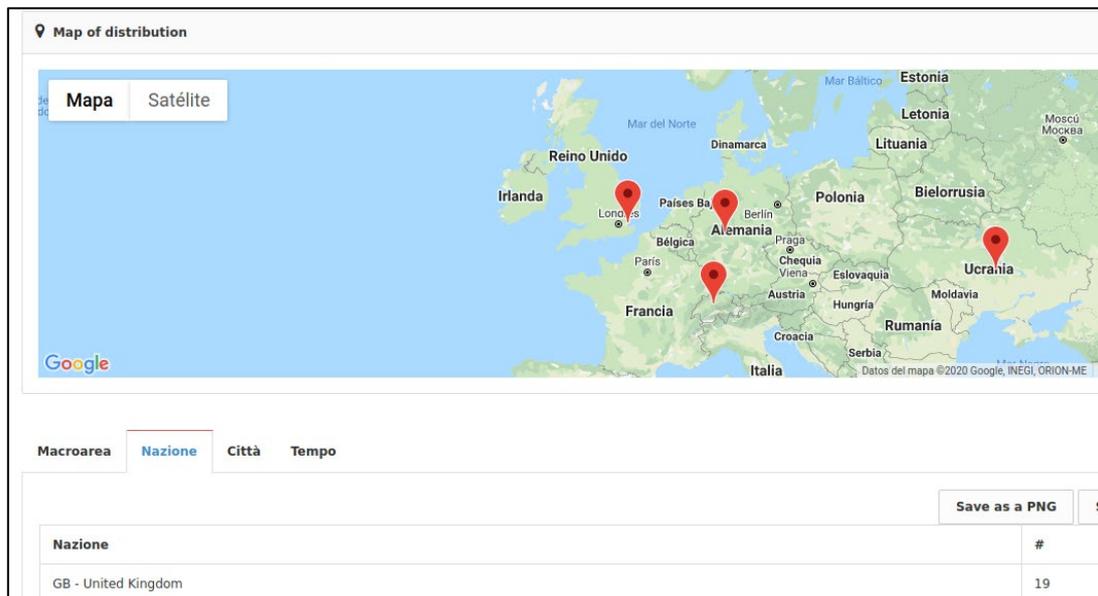


FIGURA 4. MAPA DE VISUALIZACIONES POR PAÍS DE LAS PUBLICACIONES DEL AUTOR «DE SANCTIS, UMBERTO», EN EL REPOSITORIO DE LA UNIVERSIDAD DE ESTUDIOS DE MILÁN

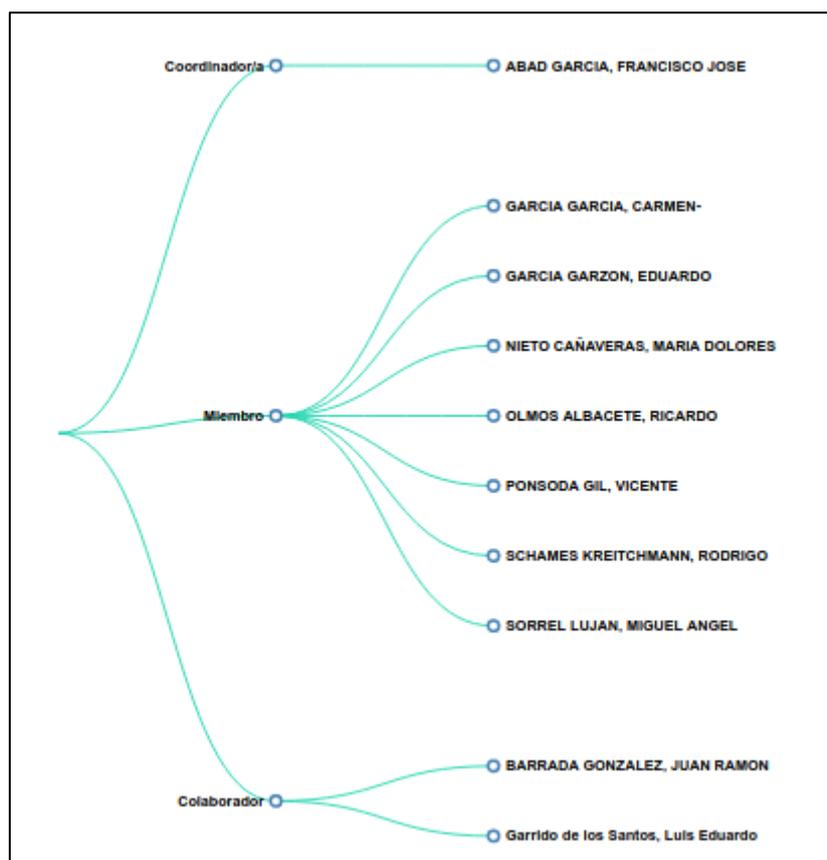


FIGURA 5. RED DE COLABORADORES DEL AUTOR «ABAD GARCÍA, FRANCISCO JOSÉ» PARA EL PROYECTO MODELOS Y APLICACIONES PSICOMÉTRICAS, DE LA UNIVERSIDAD AUTÓNOMA DE MADRID

Identificadores persistentes de autor

Un identificador persistente de autor es un registro de autor al cual se le asigna un código alfanumérico único que identifica de forma unívoca a un autor, su currículum y sus publicaciones (CORCHUELO, 2015). Los identificadores, además de permitir la desambiguación de nombres, también permiten la conexión con fuentes externas para extraer mayor información y enriquecer los datos del autor (LATIF *et al.*, 2018).

Uno de los puntos analizados fue si los perfiles de autor permiten relacionar al autor con alguno de estos identificadores. El análisis indicó que nueve de los 28 repositorios con perfiles de autor permiten la integración de identificadores persistentes en su perfil de autor. ORCID, Scopus ID y Researcher ID son los proveedores utilizados en la mayoría de ellos, e incluso uno de estos repositorios ha tomado al identificador provisto por una de estas compañías

como el identificador primario usado internamente para sus autores. Este es el caso del repositorio Europe PubMed Central, en el que la identificación de sus autores se hace a través del ORCID del investigador. En el análisis se observa que ORCID se utiliza en ocho de los 28 repositorios (28 %), Researcher ID en cinco (18 %) y Scopus ID en seis de ellos (21 %).

Cabe destacar que dentro de los 172 repositorios que no implementaron el servicio de perfil de autor de los 200 analizados, existen varios que utilizan identificadores persistentes para identificar a sus autores de forma unívoca. Por ejemplo, el repositorio de la Universidad Politécnica de Cataluña relaciona a sus autores con su ORCID, en caso de que posean uno.

Datos personales

En cuanto a los datos personales, se destaca que la mayoría de los repositorios analizados (22 de 28) muestran algún dato personal en los perfiles de autor. En este punto, se tuvieron en cuenta varios aspectos, en los que se destacan el e-mail y la foto, con un 77 % de los repositorios que los presentan en el perfil de autor; lo siguen otros aspectos como pueden ser las áreas de investigación, página web o información personal, entre otros (ver FIGURA 6).

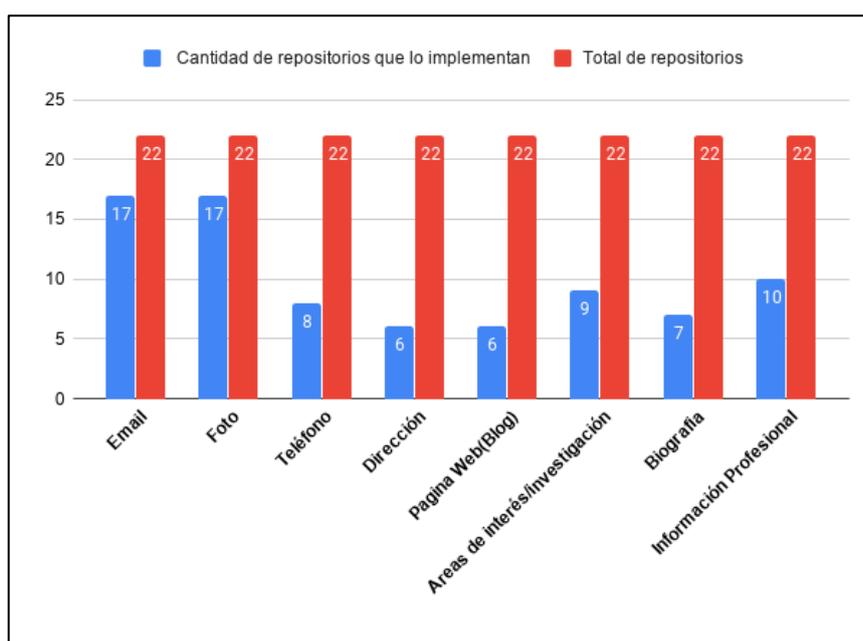


FIGURA 6. CANTIDAD DE REPOSITARIOS QUE MUESTRAN CADA UNO DE LOS DATOS PERSONALES DE LOS AUTORES

Otro dato analizado en relación a los datos personales son las redes sociales personales. En este caso, solo dos de los 28 repositorios destacan alguna red como Facebook, Twitter o LinkedIn.

Datos de la filiación

En el caso de los datos de la filiación institucional se relevó que 24 de los 28 perfiles muestran algún dato relacionado con ella. En algunos casos simplemente se muestra el nombre de la institución y en otros se encontraron descripción de la institución en detalle, con dirección (a veces integrado con mapas), teléfono y dependencias.

Redes sociales académicas

Los Academic Social Network Sites permiten organizar y visibilizar la producción científica de los autores de un modo sencillo y accesible para toda la comunidad científica nacional e internacional. Estas redes sociales ofrecen al investigador la posibilidad de disponer de un perfil propio a través del cual se puede compartir información y estar en contacto con otros académicos (QUINTAS FROUFE, 2016). Por este motivo, se consideró a las redes sociales académicas como un punto importante para el análisis de los perfiles. Se observó que siete de los 28 (25 %) repositorios muestran los enlaces a alguna red social académica, entre las que se destacan Google Scholar y ResearchGate, seguido de Publons, Dialnet y otras (ver FIGURA 7).

Como se indicó al analizar el uso de identificadores persistentes, durante el relevamiento se presentaron repositorios que, por más que no implementasen un perfil de autor, sí incorporan los enlaces a las redes sociales académicas junto al nombre del autor de la publicación.

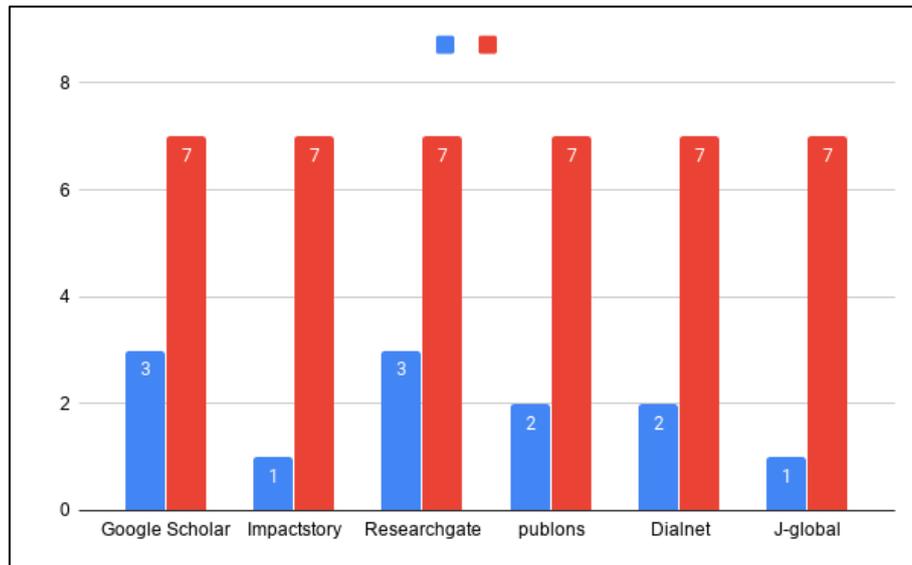


FIGURA 7. CANTIDAD DE REPOSITORIOS QUE MUESTRAN ALGÚN ENLACE A REDES SOCIALES ACADÉMICAS, AGRUPADOS POR RED SOCIAL

Resultados obtenidos

De los 200 repositorios analizados, un 14 % implementa el perfil de autor, es decir, 28 repositorios. Si se toma una muestra de los primeros 50 repositorios del ranking web de transparencia de Webometrics y se compara con la edición de 2017, se observa un incremento en la cantidad de repositorios que implementan este servicio, pasando del 10 al 20 %.

De los relevados, predomina el software DSpace con poco más del 30 %; los datos que mayormente se exponen en los perfiles tienen que ver con las publicaciones, información de la filiación y algunos datos personales (entre un 78 y 89 %), y luego siguen las estadísticas (39 %), los identificadores persistentes (32 %) y las redes sociales académicas (25 %).

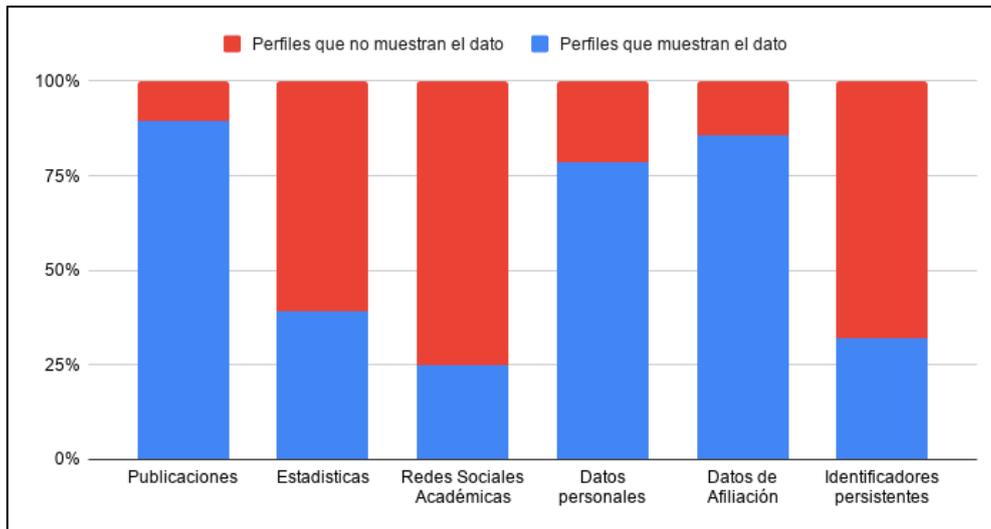


FIGURA 8. PORCENTAJE DE LOS REPOSITORIOS ANALIZADOS QUE IMPLEMENTAN CADA SERVICIO

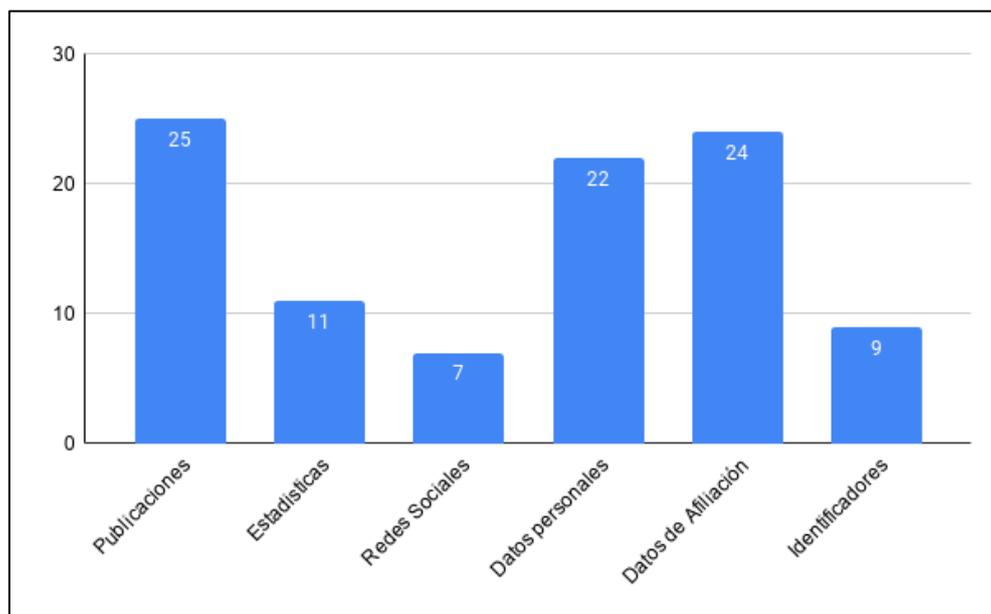


FIGURA 9. CANTIDAD DE REPOSITORIOS ANALIZADOS QUE IMPLEMENTAN CADA SERVICIO

Conclusiones

A pesar de las múltiples ventajas que conlleva la generación de perfiles de autor, la cantidad de repositorios considerados en este relevamiento que implementan este tipo de servicios continúa siendo baja. Si bien esta cantidad ha aumentado en el tiempo, no se dispone de información histórica anterior al año 2017 que permita indicar si se trata de una tendencia ascendente o un

aumento circunstancial. En caso de que en el futuro se profundice la integración de repositorios digitales con sistemas CRIS, sería posible aprovechar la implementación en formato CERIF de dichos sistemas y, de este modo, promover la adopción de servicios de perfiles de autor en los repositorios. Cabe aquí mencionar que la versión 7 del software DSpace, en etapa avanzada de desarrollo, incorpora características que promueven su integración con sistemas CRIS. Esta versión, a diferencia de las anteriores de DSpace, propone un modelo de datos más cercano al modelo de CERIF al incorporar el manejo de entidades configurables y relaciones entre ellas. Estas incorporaciones, sumadas a la adopción preferencial de DSpace entre los repositorios de la actualidad, podría acelerar la adopción de servicios de perfiles de autor en los repositorios en el corto y mediano plazo.

Bibliografía

- AGUILLO, I.; ORTEGA, J.; FERNÁNDEZ, M.; UTRILLA, A. (2010). Indicators for a webometric ranking of open access repositories. *Scientometrics*, 82, 477-486. <https://doi.org/10.1007/s11192-010-0183-y>
- BARRUECO CRUZ, J. M. y NAVALÓN, J. A. (2015). Desarrollo de perfiles de autores en DSpace para el repositorio institucional de la Universitat de València. XIV Workshop Rebiun de Poyectos Digitales / VI Jornadas OS-Repositorios (Córdoba, 11-13 de marzo de 2015). <https://helvia.uco.es/handle/10396/12622/>
- CORCHUELO, C. (2015). Identificador Persistente de Autor - Infografía ¿Qué son?, características y plataformas. [Infografía] <http://dx.doi.org/10.13140/RG.2.1.4957.0002>
- GARCÍA PEÑALVO, F. J. (2017). Cómo mejorar la visibilidad de la producción científica. El perfil del investigador [Taller]. XVIII Congreso Internacional de Investigación Educativa AIDIPE 2017, Salamanca, España. <https://repositorio.grial.eu/handle/grial/910>
- GENOVÉS, P. (2017). Perfiles de autor en repositorios institucionales. *Palabra Clave*, 7(1). <http://sedici.unlp.edu.ar/handle/10915/63375>
- LATIF, A.; BORST, T.; TOCHTERMANN, K. (2018). Compiling Scholarly Profile Pages by Integrating External Authority Data. JCDL '18: Proceedings of the 18th ACM/IEEE on Joint

Conference on Digital Libraries, pp. 411-412.

<https://doi.org/10.1145/3197026.3204473>

LORENZO ESCOLAR, N. Y PASTOR RUIZ, F. (2012). Un análisis de los principales sistemas de identificación y perfil para el personal investigador. *Aula Abierta*, 40(2), 107-118.

<https://dialnet.unirioja.es/servlet/articulo?codigo=3921021>

QUINTAS FROUFE, N. (2016). La emergencia de las redes sociales académicas: su impacto académico. *Opción*, 32(10), 517-528.

<https://dialnet.unirioja.es/servlet/articulo?codigo=5875180>

WEBOMETRICS. (n.d.). Objectives | Ranking Web of Repositories. Ranking Web of Repositories.

<https://repositories.webometrics.info/en/Objectives>

Anexo 1

El detalle de los datos recopilados se encuentra en el siguiente conjunto de datos:

<http://sedici.unlp.edu.ar/handle/10915/121507>

Pósters

Un sistema de índices para la relevancia de la producción grupal e individual

JARED DAVID TADEO GUERRERO SOSA

Universidad Autónoma de Yucatán

jared.guerrero@correo.uady.mx

VÍCTOR HUGO MENÉNDEZ DOMÍNGUEZ

Universidad Autónoma de Yucatán

mduoming@correo.uady.mx

MARÍA ENRIQUETA CASTELLANOS BOLAÑOS

Universidad Autónoma de Yucatán

enriqueta.c@correo.uady.mx

RESUMEN

Eje temático: Infraestructura tecnológica.

Descripción del problema: Los metadatos de los repositorios digitales posiblemente contienen valores erróneos, principalmente aquellos con el nombre de los autores de la publicación. Diversas instancias se encargan de evaluar a los investigadores y profesores que realizan labores de investigación con base en la producción almacenada en diversos repositorios. El objetivo del trabajo es desarrollar una solución tecnológica que valore la relevancia de un investigador a través de la evaluación de los metadatos de su producción científica, los cuales se pueden consultar en uno o más repositorios digitales.

Metodología: Se siguió la metodología de la investigación en acción, a partir de la definición del escenario de la problemática hasta la documentación y la difusión. Además, la arquitectura del sistema

propuesto se compone por capas, utilizando diferentes fuentes de datos a través de servicios web y del protocolo OAI-PMH.

Resultados: A partir del sistema propuesto se analizó la producción científica grupal e individual de 824 profesores y 78 cuerpos académicos de la Universidad Autónoma de Yucatán, México, de los cuales, se recuperó la producción de 2.858 publicaciones. A través de la interfaz del usuario es posible identificar por facultad los grupos e individuos más relevantes.

PALABRAS CLAVE

Producción científica; repositorios digitales; relevancia científica.

Scientific production, digital repositories, scientific relevance.



Un sistema de índices para la relevancia de la producción grupal e individual

Jared David Tadeo Guerrero Sosa, Víctor Hugo Menéndez Domínguez y María Enriqueta Castellanos Bolaños

Universidad Autónoma de Yucatán

{jared.guerrero, mdoming, enriqueta.c}@correo.uady.mx

Introducción

Los repositorios digitales de acceso abierto describen, localizan y hacen referencia a sus recursos por medio de los metadatos, los cuales suelen adolecer de integridad, completitud, exactitud y consistencia, principalmente aquellos que contienen el nombre de los autores de la publicación.

Diversas instancias se encargan de evaluar y reconocer a los investigadores y profesores que realizan labores de investigación con base en la producción almacenada en diversos repositorios.

El objetivo del trabajo es desarrollar una solución tecnológica que valore la relevancia de un investigador a través de la evaluación de los metadatos de su producción científica, los cuales se pueden consultar en uno o más repositorios digitales.

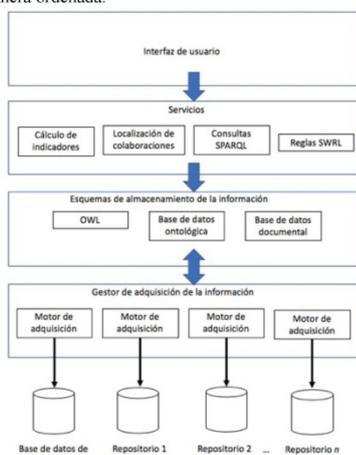
Metodología

Se siguió la metodología de la Investigación en acción (Avison, 1999) con las siguientes fases:

- Definir el escenario de la problemática y análisis del estado de arte.
- Proponer las herramientas y las métricas que ayudarán a la solución de la problemática planteada.
- Implementar un prototipo para solucionar la problemática.
- Realización de pruebas.
- Documentación y difusión.

Implementación

La arquitectura se ha diseñado en capas, de manera ordenada y facilitando el flujo de la información entre ellas y de manera ordenada.



Fuente: Guerrero Sosa (2019)

Las fuentes de datos utilizadas en la propuesta son Scopus, cuya recuperación de metadatos se ejecuta por diversos servicios web, y el repositorio nacional de México, que permite el consumo de sus metadatos a través del protocolo para la interoperabilidad OAI-PMH.

Resultados

A partir del sistema propuesto se analizó la producción científica grupal e individual de 824 profesores y 78 cuerpos académicos de la Universidad Autónoma de Yucatán, México, de los cuales, se recuperó la producción de 2,858 publicaciones. A través de la interfaz del usuario es posible identificar por facultad los grupos e individuos más relevantes.



Fuente: Guerrero-Sosa (2021)



Fuente: Guerrero-Sosa (2021)

Conclusiones

Se presentó un sistema que permite evaluar la relevancia de la investigación científica a partir de diferentes repositorios, considerando las posibles inconsistencias en los metadatos. La arquitectura propuesta, al no depender de una herramienta tecnológica específica permite añadir nuevas fuentes de datos y operar de manera ordenada.

Referencias

- Avison, D., Lau, F., Myers, M. y Nielsen, P. A. (1999). Action research. *Communications of the ACM*, 42(1), 94–97. <https://doi.org/10.1145/291469.291479>
- Guerrero-Sosa, J. D. T., Menéndez-Domínguez, V. H. y Castellanos-Bolaños, M. E. (2021). An indexing system for the relevance of academic production and research from digital repositories and metadata. *The Electronic Library*, 39(1), 33–58. <https://doi.org/10.1108/EL-06-2020-0160>
- Guerrero Sosa, J. D. T., Menéndez Domínguez, V. H., Castellanos Bolaños, M. E. y Gómez Montalvo, J. R. (2019). Use of an Ontological Model to Assess the Relevance of Scientific Production. *IEEE Latin America Transactions*, 17(9), 1424–1431. <https://doi.org/10.1109/TLA.2019.8931135>

Evaluación de los repositorios institucionales del sector agrario y afines en el Perú: estudio preliminar

SILVANA AQUINO REMIGIO

Universidad Nacional Mayor de San Marcos

silvana.aquino@gmail.com

RESUMEN

Eje temático: Evaluación y métricas alternativas.

La sistematización de la información especializada es un eje prioritario para el sector agrario y afines en el Perú, por ello es importante contar con repositorios digitales. Este estudio preliminar tiene como objetivo evaluar los cinco repositorios institucionales que constituyen este sector, a fin de identificar su estado actual y brindar recomendaciones de mejora. Para llevar a cabo el procedimiento de evaluación es necesario utilizar un conjunto de indicadores claros y comprobables; en este sentido, la lista de verificación de criterios para la evaluación contenida en el Anexo 1 de la Guía ALICIA (documento normativo de los repositorios institucionales peruanos), se constituye como una herramienta adecuada para este fin. La metodología utilizada comprende un análisis exploratorio y la aplicación de la mencionada lista de verificación, que cuenta con 55 criterios agrupados en siete secciones: a) visibilidad, b) políticas, c) aspectos legales, d) metadatos, e) interoperabilidad, f) *logs* y estadísticas y g) seguridad, autenticidad e integridad de los datos. Se presentarán resultados y recomendaciones para cada uno de los criterios mencionados. Estos resultados muestran que estos repositorios cumplen parcialmente los criterios establecidos en la Guía ALICIA y vienen mejorando gradualmente.

PALABRAS CLAVE

Repositorios institucionales; acceso abierto; sector agrario y afines.

Institutional Repositories; Open Access; Agricultural and related sector.



BIREDIAL-ISTEC 2021

X Conferencia Internacional Bibliotecas y Repositorios Digitales (virtual)

Evaluación de los repositorios institucionales del sector agrario y afines en el Perú: estudio preliminar

Silvana Aquino-Remigio*

*Universidad Nacional Mayor de San Marcos, Lima, Perú

INTRODUCCIÓN

Para el sector agrario y afines en Perú es importante contar con repositorios digitales que sistematicen su información especializada de acuerdo a criterios estandarizados, como los que contiene el Anexo 1 de la Guía ALICIA (documento normativo de los Repositorios Institucionales - RI adheridos al Repositorio Nacional ALICIA).

OBJETIVO

Evaluar los RI del sector agrario y afines del Perú, a fin de identificar su estado actual y brindar recomendaciones de mejora.

Repositorios Institucionales (RI) evaluados



MÉTODO

Comprende un análisis exploratorio y la aplicación de una lista de verificación, con criterios agrupados en 7 secciones: 1) visibilidad, 2) políticas, 3) aspectos legales, 4) metadatos, 5) interoperabilidad, 6) logs y estadísticas y 7) seguridad, autenticidad e integridad de los datos.

RESULTADOS Y RECOMENDACIONES

- 1) **Visibilidad:** Todos tienen enlace desde plataforma Gob.pe. La mayoría tiene presencia en directorios y recolectores. Carecen de adhesiones a declaraciones internacionales a favor del AA. En términos generales todos estos RI vienen mejorando paulatinamente su visibilidad. Se recomienda reforzar iniciativas para fomentar visibilidad dentro de sus instituciones.
- 2) **Políticas:** 3 RI cuentan con documento. Se recomienda procurar que éste sea emitido a través de una resolución (ejecutiva, directoral, etc.) según corresponda.
- 3) **Aspectos legales:** Usan licencias *Creative Commons (CC) 4.0* en registros. Se recomienda mejorar orientación para depositar documentos.
- 4) **Metadatos:** Usan metadatos obligatorios para ser cosechados por ALICIA. Se recomienda mejorar calidad de metadatos, pues esta es irregular.
- 5) **Interoperabilidad:** Usan DSpace y protocolo OAI-PMH. Se recomienda configurar actualización de la cosecha.
- 6) **Logs y estadísticas:** Solo 2, comparten de manera pública las estadísticas. Se recomienda gestionar configuraciones y transparentar los datos estadísticos.
- 7) **Seguridad, autenticidad e integridad de los datos:** Realizan copias de seguridad correspondientes. Se recomienda incluir dentro de las políticas este aspecto.

Resultados de evaluación

RI	PROMEDIO GLOBAL DE EVALUACIÓN
[R1] Repositorio MIDAGRI	66.67%
[R2] Repositorio ANA	95%
[R3] Repositorio INIA	87%
[R4] Repositorio SENASA	58.97%
[R5] Repositorio SERFOR	90%

Nota: Datos de agosto 2021

CONCLUSIÓN

Los RI del sector agrario y afines cumplen con criterios establecidos en la Guía ALICIA y vienen mejorando gradualmente. Ello demuestra su compromiso con el acceso abierto y su interés por visibilizar la producción e investigación de las instituciones que representan. Cabe señalar, que cuentan también con el acompañamiento del CONCYTEC.

REFERENCIAS

CONCYTEC (2021). Guía ALICIA 2.0.1: Directrices para repositorios institucionales de la Red Nacional de Repositorios Digitales de Ciencia, Tecnología e Innovación de Acceso Abierto (RENARE). <http://repositorio.concytec.gob.pe/handle/20.500.12390/2231>



Representación de la colaboración de cuerpos académicos por medio de la teoría de grafos

JARED DAVID TADEO GUERRERO SOSA

Universidad Autónoma de Yucatán

jared.guerrero@correo.uady.mx

MARÍA ENRIQUETA CASTELLANOS BOLAÑOS

Universidad Autónoma de Yucatán

enriqueta.c@correo.uady.mx

VÍCTOR HUGO MENÉNDEZ DOMÍNGUEZ

Universidad Autónoma de Yucatán

mdoming@correo.uady.mx

RESUMEN

Eje temático: Evaluación y métricas alternativas.

Descripción del problema: En México existen grupos de investigación conocidos como cuerpos académicos, los cuales pertenecen a una institución y su composición consiste en profesores de tiempo completo. A través de evaluaciones, las cuales pueden ser subjetivas por la falta de herramientas para dicho proceso, se conoce el impacto de su producción y son clasificados de acuerdo con el grado de consolidación. El objetivo del trabajo es proponer el análisis de cuerpos académicos a partir de los metadatos de la producción indizada por Scopus, haciendo uso de la visualización de datos a través de la teoría de grafos.

Metodología: Se siguió la metodología de la minería de datos para aplicaciones e-learning, considerando la recolección de datos, el procesamiento de datos para el cálculo de indicadores de colaboración y contribución y la visualización por medio de grafos.

Resultados: Se analizó la producción científica en Scopus de tres cuerpos académicos de la Universidad Autónoma de Yucatán, México, de los cuales se calcularon los indicadores y se generaron los grafos, identificando que mientras mayor sea el grado de consolidación, menor será el porcentaje de contribución promedio.

PALABRAS CLAVES

Colaboración científica; cuerpos académicos; producción científica; teoría de grafos; visualización de datos.

Scientific collaboration; institutional research groups; scientific production; graph theory; data visualization.



Representación de la colaboración de cuerpos académicos por medio de la teoría de grafos

Jared David Tadeo Guerrero Sosa, María Enriqueta Castellanos Bolaños y Víctor Hugo Menéndez Domínguez
 Universidad Autónoma de Yucatán
 {jared.guerrero, enriqueta.c. mdoming}@correo.uady.mx

Introducción

En México existen grupos de investigación conocidos como cuerpos académicos, los cuales pertenecen a una institución y su composición consiste en profesores de tiempo completo.

A través de evaluaciones, las cuales pueden ser subjetivas por la falta de herramientas para dicho proceso, se conoce el impacto de su producción y son clasificados de acuerdo con el grado de consolidación.

En este trabajo se propone el análisis de cuerpos académicos a partir de los metadatos de la producción indexada por Scopus, haciendo uso de la visualización de datos a través de la teoría de grafos.

Metodología

Se siguió la metodología de la minería de datos para aplicaciones de e-learning (Prieto, 2011) por medio de las siguientes fases:

Recolección de datos: Se obtiene la información de los integrantes de los cuerpos académicos y su respectiva producción en Scopus.

Procesamiento de datos: Consiste en depurar publicaciones y calcular el grado de colaboración y el porcentaje de contribución de cada integrante en su cuerpo académico.

El grado de colaboración de un integrante en cada publicación se obtiene por medio de la fórmula:

$$GC = 2 * 2^{-p}$$

donde p es la posición del integrante en la lista de autoría. El porcentaje de contribución de un integrante en su cuerpo académico se obtiene por medio de la fórmula

$$PC = (100 * \sum_{i=1}^n GC_i) / t$$

donde:

n es el número de publicaciones en las que participa el integrante del cuerpo académico

GC es el grado de colaboración en la i -ésima publicación

t es el total de publicaciones del cuerpo académico

Visualización: A partir de los resultados de las fases previas, construir las tablas de los grafos para ser representados a través del software NodeXL.

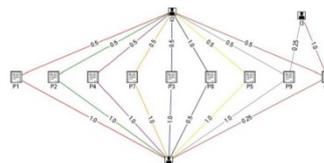
Caso de estudio

Se analizó la producción en Scopus hasta el mes de agosto de 2019 de tres cuerpos académicos de la Universidad Autónoma de Yucatán, cada uno de diferente grado de consolidación.

Los vértices representan a los integrantes de los cuerpos académicos y sus publicaciones y el peso de cada arista indica el grado de colaboración de cada integrante en cada publicación.

Cuerpo académico en formación.

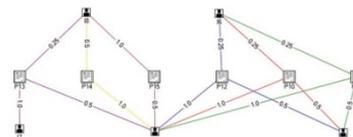
- Número de integrantes: 3
- Número de publicaciones: 9
- Porcentaje de contribución promedio: 51.83%



Fuente: Menéndez Domínguez, 2020

Cuerpo académico en consolidación.

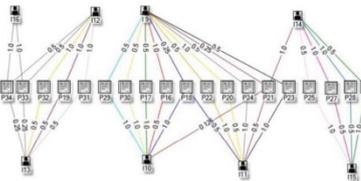
- Número de integrantes: 5
- Número de publicaciones: 6
- Porcentaje de contribución promedio: 33.3%



Fuente: Menéndez Domínguez, 2020

Cuerpo académico consolidado.

- Número de integrantes: 8
- Número de publicaciones: 19
- Porcentaje de contribución promedio: 19.48%



Fuente: Menéndez Domínguez, 2020

Conclusiones

Se presentó una evaluación de cuerpos académicos a través de la teoría de grafos, la cual permite identificar los elementos con mayor aportación gracias a los indicadores propuestos. Se analizaron tres cuerpos académicos de diferente consolidación, obteniendo tres escenarios diferentes.

Referencias

Menéndez Domínguez, V. H., Guerrero Sosa, J. D. T., Castellanos Bolaños, M. E. y Zurita Gallegos, E. (2020). Análisis de la producción de cuerpos académicos basado en teoría de grafos. RIDE Revista Iberoamericana Para La Investigación Y El Desarrollo Educativo, 10(20). <https://doi.org/10.23913/ride.v10i20.603>

Prieto, M. E., Zapata, A. y Menéndez, V. H. (2011). Data Mining Learning Objects. En C. Romero, S. Ventura, M. Pechenizkiy y R. Baker (Eds.), Handbook of Educational Data Mining (pp. 481–492). CRC Press.

Ampliação da ciência aberta na Universidade Federal de Minas Gerais: ações e perspectivas para ampla visibilidade dos resultados de pesquisa

IZABEL ANTONINA DE ARAÚJO

Universidade Federal de Minas Gerais

antoninaizabel@gmail.com

CAMILA MARIANA APARECIDA DA SILVA

Universidade Federal de Minas Gerais

camilamas.silva@gmail.com

RESUMO

Eixo temático: Comunicação acadêmica e científica aberta.

Garantir infraestrutura tecnológica para difundir a produção científica é condição fundamental para ampliação da ciência aberta. O artigo aborda iniciativas da Universidade Federal de Minas Gerais na promoção do acesso aberto à sua produção intelectual. A instituição trabalha para ampliar a visibilidade dos resultados de pesquisa visando promover a acessibilidade e a difusão dos resultados de pesquisa que se aproximam de 40 mil itens documentais produzidos anualmente. Trata-se de pesquisa descritiva e aplicada que utilizou como objeto de investigação sistemas informacionais de disponibilização da produção intelectual da UFMG. Descreve iniciativa para criar processos facilitadores para integrar o repositório a uma base de dados - Importa Lattes - e aos sistemas acadêmicos da universidade. Contextualiza o Repositório Institucional da UFMG, as iniciativas de importar os dados referenciais padronizados da Plataforma Lattes e exportá-los para o Repositório, a proposta de integração dos sistemas informacionais e a síntese dos resultados

alcançados até o momento. Conclui-se que a importação dos dados de forma padronizada para o repositório facilita a gestão e agiliza a disponibilização da informação. Ressalta-se que para garantir celeridade na disponibilidade dos resultados e o gerenciamento eficaz da produção científica, são necessários investimentos em infraestrutura informacional com integração e unificação dos processos de trabalho.

PALAVRAS-CHAVE

Ciência aberta; produção científica; integração entre sistemas de informação.

ABSTRACT

Ensuring technological infrastructure to disseminate scientific production is a fundamental condition for expanding open science. The article addresses the initiatives of the Federal University of Minas Gerais in promoting open access to its intellectual production. The institution works to increase the visibility of academic research results to promote accessibility and dissemination of research results that approach 40,000 total documentary items. This is a descriptive and applied research that uses as its object of investigation informational systems for making available the intellectual production of UFMG. Describes an initiative to create facilitating processes to integrate the repository to a database - Importa Lattes - and to the university's academic systems. It contextualizes the UFMG Institutional Repository, as an initiative to import standardized referential data from the Lattes Platform and export them to the Repository, a proposal for the integration of informational systems and a synthesis of the results achieved so far. Conclude that importing data in a standardized way to the repository facilitates management and speeds up the availability of information. It is noteworthy that to ensure speed in the availability of results and effective management of scientific production, investments are made in informational infrastructure with integration and unification of work processes.

KEYWORDS

Open science; scientific production; integration between information systems.

Introdução

A ciência aberta, compreendida por MARTINEZ FLUENTE e VICENTE SÁEZ (2018) como um termo guarda-chuva que busca promover a remoção das barreiras de compartilhamento de resultados, métodos e ferramentas em qualquer fase do processo de pesquisa, representa novos desafios e perspectivas para comunicação e difusão dos trabalhos científicos produzidos nas universidades. Requer pensar em novos modelos interativos e colaborativos para disseminar e usar o conhecimento produzido da forma mais rápida e ampla possível, tendo em vista que a difusão dos resultados de pesquisa é essencial nas práticas científicas. Neste contexto de abertura, visando ser reconhecidas pela transparência pública, contribuir para o avanço do conhecimento científico e impacto social, universidades de países dos cinco continentes vêm intensificando as ações para oferecer maior grau de abertura aos resultados de pesquisa produzidos. Conforme anunciam CABALLERO RIVERO, SÁNCHEZ TARRAGÓ, SANTOS (2019), a abertura do conhecimento, além de promover o impacto social da ciência, fomenta a reprodutibilidade dos resultados de pesquisa, a transparência metodológica, o acesso global aos resultados de investigação e a economia de recursos (tempo, dinheiro, equipe, equipamento).

A construção de robustas infraestruturas institucionais e nacionais é essencial para dar visibilidade aos resultados de pesquisa produzidos no Brasil. Conforme GODIN e DORÉ (2005), os aspectos técnicos e tecnológicos estão relacionados à infraestrutura, compreendendo aspectos relacionados aos softwares e serviços de informação para promover inovação. De igual modo, é essencial para o avanço da sociedade, a orientação no sentido de uma

educação transformadora, com a proposição de novos currículos, criação de ferramentas pedagógicas, incentivo à qualificação, desenvolvimento de competências em pesquisas, bem como adequação entre formação e trabalho e utilização de conhecimentos adquiridos.

Nesta perspectiva, o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), promove o desenvolvimento de recursos e a infraestrutura de informação em ciência e tecnologia para a produção, socialização e integração do conhecimento científico e tecnológico. Por meio de um repositório coletor da produção científica, o Oasisbr, coleta a produção das universidades e instituições de pesquisas e exporta os dados para bases de dados internacionais como a *La Referencia* e a *Confederation of Open Access Repositories* (COAR). Todavia, esta ação pode alcançar resultados mais amplos se contar com a parceria das universidades em promover alternativas locais de infraestrutura de sistemas de divulgação científica. Investimentos desta ordem apresentam-se como desafio tendo em vista que em muitas instituições não há um sistema unificado e integrado para disponibilizar a produção intelectual. Para informar os dados e resultados de pesquisa, a fim de cumprir as exigências institucionais, de fundações que regulam a pós-graduação e de agências de fomento, docentes e pesquisadores precisam fornecer as mesmas informações a vários sistemas diferentes. Considerando a ocupação com uma gama de atividades diárias, essa tarefa nem sempre é priorizada nas atividades desses profissionais, o que resulta em diminuição significativa da visibilidade potencial, pois estes conteúdos muitas vezes não são disponibilizados nos repositórios locais. Portanto, tampouco são coletados pelo IBICT.

A Universidade Federal de Minas Gerais (UFMG) se insere no movimento de abertura do conhecimento científico, buscando alternativas para povoar seu Repositório Institucional (RI-UFMG), através da criação de políticas de informação e iniciativas para implementação de projetos para automatizar processos e integrar sistemas de divulgação do conhecimento científico. Esta inserção é imprescindível, uma vez que a Universidade se consolida como uma importante produtora de conhecimento de interesse público em todas as áreas do conhecimento.

O RI-UFMG foi implementado em 2019 para integrar conteúdos relativos à produção intelectual dos docentes, discentes e técnico-administrativos em educação, compreendendo artigos de periódicos e de eventos, livros, capítulos de livros, teses, dissertações, monografias, trabalhos de conclusão de curso, entre outros, abrangendo mais de 40 mil documentos anuais. Desde a implantação do sistema, a Universidade vem buscando meios para dar maior agilidade ao povoamento do Repositório. Para tanto, tem investido em alternativas para importar a produção retrospectiva da Plataforma Lattes para promover a integração entre os sistemas de informação e o repositório, tendo como propósito agilizar o processo de disponibilização da produção acadêmica.

O presente artigo descreve iniciativas para ampliar o acesso aberto à produção intelectual da UFMG através de estudos realizados com o intuito de integrar sistemas informacionais e a importação de dados referenciais padronizados da Plataforma Lattes para o Repositório. Com a realização de uma pesquisa descritiva e aplicada investigou-se os sistemas informacionais de disponibilização da produção intelectual e as práticas de disponibilização da informação a fim de propor um modelo alternativo para ampliar a divulgação dos resultados de pesquisa. Como resultado, apresenta as vantagens de integração dos sistemas e a implementação de uma base de dados integrada ao Repositório que permite realizar o gerenciamento dos 78 mil itens importados da Plataforma Lattes.

O estudo, além de trazer reflexão para esse novo cenário de mudança do processo de divulgação da pesquisa, busca fomentar o debate sobre a necessidade de estabelecer infraestruturas com sistemas integrados e interoperáveis para facilitar o gerenciamento da informação científica.

Procedimentos metodológicos

Com base na revisão de literatura em experiências relatadas por MATIAS (2015), VIDOTTI, BASTOS, GRISOTO, ARAKAKI e FERREIRA (2016) e SARVO, FARIA, MATIAS, MATIAS e AMARAL (2016) sobre a importação de dados a partir da Plataforma

Lattes e no conhecimento das pesquisadoras na elaboração de iniciativas de acesso de aberto à produção científica, foi realizado um estudo para entender os processos informacionais em vigência na instituição, visando identificar pontos de melhorias para otimizar processo de trabalho e ampliar a visibilidade da produção intelectual da UFMG. Trata-se de uma pesquisa descritiva e aplicada que utilizou como objeto de investigação os sistemas informacionais de disponibilização da produção intelectual da UFMG e práticas a eles associadas. O estudo constou de três etapas:

- Avaliação dos fluxos informacionais realizados nos sistemas de disponibilização da informação acadêmica e científica;
- Análise de alternativas para povoar o Repositório Institucional a partir de produção importada da Plataforma Lattes;
- Elaboração de propostas para integrar sistemas para ampliar a visibilidade dos produtos de pesquisas produzidos na instituição.

Como técnica para coleta de dados, utilizou-se o método qualitativo através da técnica de observação e acompanhamento dos processos informacionais e dos fluxos de trabalho vigentes.

Na etapa inicial identificou-se que os processos vigentes para disponibilização da produção caracterizavam-se pela morosidade e resultavam em dificuldades de execução por parte dos usuários em função da utilização de sistemas de informação distintos e sem nenhuma conexão entre si. Observou-se também a necessidade de criar estrutura interoperável de sistemas, redesenho dos fluxos correntes, bem como a criação de novo fluxo informacional para atendimento à demanda do MEC/INEP para disponibilizar os Trabalhos de Conclusão de Curso (TCC) de graduação.

Na segunda etapa, constatou-se que para dar celeridade ao processo de povoamento das publicações científicas, seria necessária a construção de uma base integrada ao software DSpace para armazenamento dos dados

referenciais importados da plataforma Lattes, haja vista a expressividade do número de itens recuperados na importação.

A última etapa resultou nas propostas de integração de sistemas, aprimoramento dos fluxos informacionais e implantação da base integrada ao Repositório, conforme discussão no próximo tópico.

Proposta de integração de sistemas de divulgação da produção científica: iniciativas para ampliar a visibilidade dos resultados de pesquisas produzidos na UFMG

A ampliação da Ciência aberta requer maior celeridade na disponibilização da informação. Faz-se necessária a adoção de sistemas acadêmicos integrados e interoperáveis com os repositórios institucionais a fim de viabilizar a ampla visibilidade da produção científica. Nesta perspectiva, a equipe do RI-UFMG apresentou duas propostas visando unificar e automatizar os processos para o depósito e divulgação de trabalhos acadêmicos (teses dissertações, monografias e trabalhos de conclusão de curso) e dar celeridade à divulgação da produção científica importada da Plataforma Lattes (livros, capítulos de livros, artigos de periódico e de evento), ampliando a comunicação com os docentes.

Proposta 1	Projeto SIGA: celeridade em disponibilização de trabalhos acadêmicos e emissão de documentação associada
Proposta 2	Importação dos dados referenciais da produção científica e a construção de uma base paralela integrada ao Repositório para armazenar e dar acesso aos itens importados da plataforma Lattes.

Fonte: elaborado pelas autoras, 2021

Projeto integração do sistema de gestão acadêmica com o Repositório

A implementação do projeto SIGA visa garantir automação e unificação do processo de autoarquivamento da versão digital dos trabalhos finais de pós-

graduação, aproveitando a infraestrutura de sistema já existente na instituição e propondo sua customização para atendimento à demanda de fluxo informacional e interoperabilidade com o repositório institucional. Por meio deste projeto, todos os passos, desde a marcação da defesa até a disponibilização das teses, dissertações e monografias, serão realizados em um único sistema, superando a divisão hoje vigente devido ao uso concomitante de dois sistemas paralelos. Isto dará agilidade à disponibilização dos trabalhos para consulta pública e emissão do atestado de entrega necessário à solicitação de diploma pelos egressos, além de permitir a descentralização do fluxo, a cargo da equipe do repositório, para as bibliotecas das unidades acadêmicas mediante criação de diferentes perfis de acesso. A concepção geral do fluxo atual e do fluxo proposto pode ser observada na FIGURA 1 abaixo:

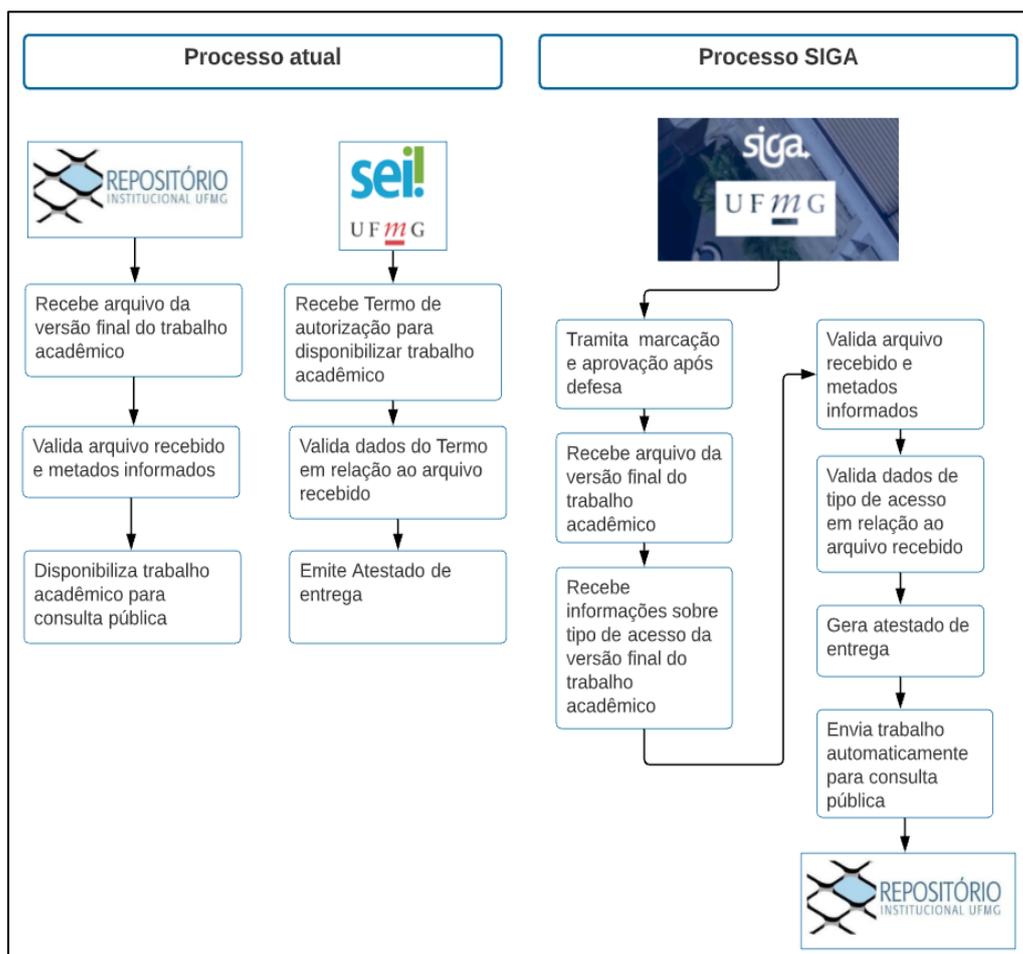


FIGURA 1. FLUXO DE INFORMAÇÃO ATUAL E PROPOSTO

Fonte: elaborado pelas autoras, 2021

Conforme ilustra a FIGURA 1, o processo atual é centralizado no Repositório, requer ações em sistemas não integrados, demanda esforços dos autores envolvidos no processo para localizar e validar a informação. Além de demandar gerenciamento e curadoria de dados em duplicidade, ocasionando erros e morosidade na disponibilização do trabalho no repositório.

Na proposta com a integração dos sistemas, todos os atores envolvidos (discentes, orientadores, secretarias, equipe do repositório, equipe do sistema de bibliotecas e Departamento de Registro e Controle Acadêmico (DRCA) visualizam e executam as tarefas atribuídas dentro do Sistema Integrado de Gestão Acadêmica (SIGA) a partir do perfil de usuário correspondente. O discente deposita a versão final aprovada pela banca e nas bibliotecas das unidades acadêmicas é feita a conferência do arquivo entregue relativamente à sua adequação às normativas institucionais, havendo possibilidade de comunicação com o discente nos casos em que alterações no arquivo se façam necessárias. No que concerne aos orientadores, será possível certificar que o arquivo depositado corresponde à versão final do trabalho consolidada sob sua supervisão, bem como conferência quanto ao tipo de acesso facultado ao arquivo.

Posteriormente à validação dos envolvidos, às secretarias e DRCA é viabilizada a extração de documentação necessária ao prosseguimento do processo de solicitação de diploma. O módulo criado pelo projeto SIGA é espelho dos metadados utilizados no repositório institucional, permitindo o fornecimento de todos os dados necessários e sua exportação para o repositório. Os metadados são automaticamente preenchidos no DSpace, o arquivo é anexado e o trabalho fica disponível para consulta pública conforme permissões de acesso concedidas pelo autor.

As alterações simplificam o processo na medida em que passam a ocorrer em um único sistema, agilizam a disponibilização dos trabalhos para consulta pública e emissão de documentação associada, além de conferirem maior confiabilidade ao processo pelo encadeamento de todas as etapas e atores envolvidos.

A implantação do projeto SIGA terá reflexos diretos sobre a graduação, uma vez que será usado como parâmetro para o desenvolvimento de um novo módulo que permitirá tramitar os trabalhos de conclusão de curso de graduação (TCCs). Em atendimento ao «Instrumento de Avaliação de Cursos de Graduação Presencial e a Distância», do MEC/INEP, a atribuição de nota 5 aos cursos está condicionada à disponibilização destes trabalhos em Repositório Institucional. Deste modo, é imperativa a extensão das ações do repositório no tocante aos cursos de graduação.

A fim o cumprimento das ações pleiteadas com o Projeto SIGA demanda, também, mobilização de ações no sentido de buscar atualização das normas relativas à entrega da versão digital dos trabalhos no âmbito dos cursos de pós-graduação, em conformidade com a Resolução CEPE nº 11 de 2019, que cria o Repositório Institucional da UFMG e estabelece sua política informacional.

Criação de uma base integrada ao repositório

Com o propósito de ampliar a visibilidade dos resultados de pesquisa produzidos na Universidade, foi trabalhado o arquivamento mediado através da importação das produções retrospectivas da Plataforma Lattes para o Dspace. Inicialmente foram importados mais de 40 mil itens com dados referenciais da produção correspondente ao ano de 2016.

Após a importação foi realizado um processo de separação, desambiguação e associação de itens com preenchimento automático dos metadados conforme padrão Dublin Core. Para maior qualidade da informação foram definidos filtros automáticos e regras de importação de modo a garantir que a forma de preenchimento dos dados esteja em conformidade com as diretrizes do IBICT e da base de dados La Referência. Este procedimento facilita o trabalho de validação dos bibliotecários, tendo em vista que o preenchimento dos dados na Plataforma Lattes não é normalizado e as regras aplicadas asseguram padronização mínima. O ajuste manual de todos dados seria bastante oneroso e colocaria em questão a efetividade da importação.

No entanto, o Dspace não suportou o armazenamento de um número tão expressivo de itens em uma única fila para validação dos dados, havendo problemas para o carregamento do sistema, especialmente quando ocorriam vários acessos simultâneos. Desde modo, a solução encaminhou-se para a criação de uma base, denominada Importa Lattes, que permite o armazenamento de todos itens importados posteriormente à aplicação das regras visando à melhoria na qualidade dos dados colhidos. O envio de itens para a fila de validação do Repositório passa a ser feito por lotes, diminuindo substancialmente a sobrecarga no DSpace. A base conta com os metadados principais de identificação como campo de busca da produção importada, além de permitir a associação de filtros que aumentam a precisão na recuperação de itens a serem enviados para o Repositório.

Resultados parciais

O projeto SIGA encontra-se em andamento e foi considerado inovador pelos órgãos administrativos da universidade, tendo sido solicitada prioridade em seu desenvolvimento em consideração ao alto nível de impacto em diferentes setores da instituição e por implicar em racionalização do fluxo vigente. Além da economia de recursos por evitar duplicidade de trabalhos e agilizar o processo de disponibilizar os trabalhos acadêmicos no repositório.

A base integrada ao RI-UFMG, possibilitou a importação de 78 mil dos dados referenciais da produção relativa aos anos de 2016 a 2019. Por meio da base Importa Lattes é possível acessar a produção importada. Os dados podem ser visualizados por meio de filtros de busca (autor, título, ano, departamento, tipo de documento). Os itens são selecionados e exportados para o repositório já em associação automática com as coleções pertinentes (artigo de periódico, artigo de evento, livro, capítulo de livro). Em caso de itens importados em duplicidade, a base permite a exclusão, melhorando os mecanismos de gestão da fila de validação do Repositório e na própria base. Ademais, é possível realizar buscas na base pelos itens já importados para o Repositório, pelos não importados e, também, dentre os excluídos.

A Figura 2 ilustra o processo de armazenamento e exportação dos dados referenciais importados da plataforma Lattes para o RI-UFMG e como os dados podem ser recuperados pela base.

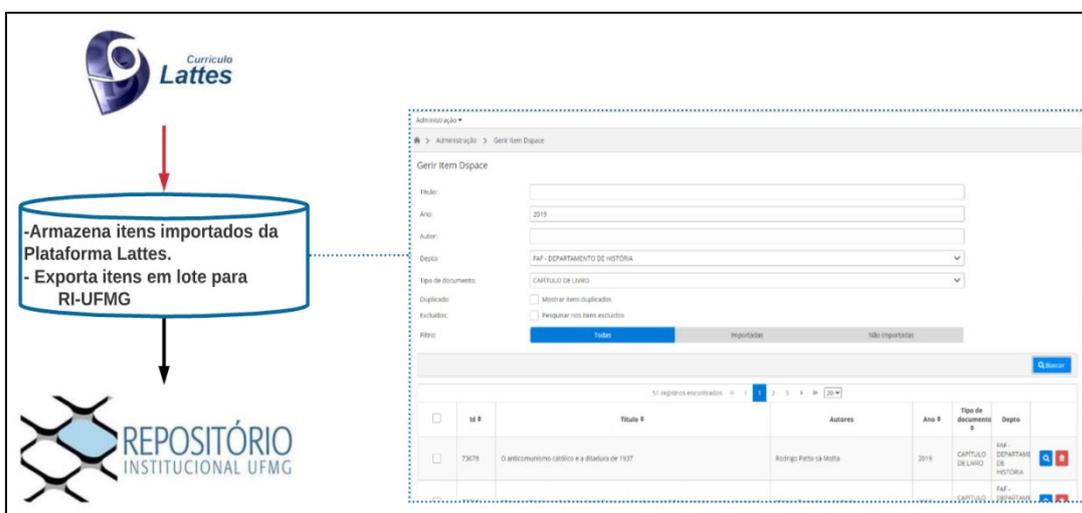


FIGURA 2. BASE IMPORTA LATTES INTEGRADA AO REPOSITÓRIO

Fonte: elaborado pelas autoras, 2021

Para garantir a qualidade da informação, antes de sua disponibilização, a produção é tratada pelos bibliotecários do sistema de bibliotecas mediante conferência do preenchimento dos metadados com base nos textos completos dos itens importados. Um dos aspectos mais relevantes neste processo consiste na verificação do tipo de acesso imposto pelo publicador. Dada a ausência deste metadado nos itens extraídos no Lattes, faz-se necessária conferência da informação anteriormente à disponibilização do item.

Considerando a importância de campanhas de sensibilização para comunidade acadêmica frente ao pouco conhecimento sobre o movimento da ciência aberta e a necessidade de estreitamento de laços entre o Repositório e os pesquisadores objetivando adesão e a promoção de confiabilidade do serviço, foram encaminhadas comunicações pessoais individualizadas para os docentes da instituição. Nestas comunicações, os docentes foram notificados sobre a iniciativa de disponibilização dos resultados de pesquisa pelo Repositório Institucional, sendo explicadas as questões de respeito às

restrições editoriais impostas pelos publicadores em conformidade com a normativa institucional vigente que regula a questão - Resolução CEPE, nº 11 de 2019. Em 70% dos envios a equipe obteve retorno dos docentes no sentido de demonstrar apoio e corroborar a relevância da iniciativa.

Resultados finais

Considerando as iniciativas para ampliação das ações voltadas ao estímulo à ciência aberta na UFMG através do uso dos dados do Currículo Lattes para mediar o povoamento do Repositório Institucional para as publicações e a unificação de sistemas para garantir celeridade e confiabilidade do autoarquivamento de trabalhos acadêmicos, depreende-se que as propostas apresentadas possuem grande potencial para o gerenciamento mais efetivo da produção científica. Nesta perspectiva os resultados finais pretendidos são:

- otimizar os fluxos de trabalho pela unificação de etapas em sistema único;
- promover a segurança da informação;
- disponibilizar 78 mil itens importados;
- ampliar a visibilidade dos resultados de pesquisa;
- estreitar comunicação e parceria com os pesquisadores da universidade;
- ampliar a cultura do acesso aberto à produção científica na Universidade;
- adequar a universidade às normas do MEC/INEP para disponibilizar os TCCs.

Constata-se com o desenvolvimento do estudo, que diante das novas tendências informacionais e científicas para ampliar a visibilidade dos produtos de pesquisa, para garantir celeridade na disponibilidade dos resultados e o gerenciamento eficaz da produção científica, são necessários investimentos

em infraestrutura informacionais com unificação dos processos. Ressalta-se a importância de campanhas de sensibilização junto à comunidade acadêmica para conhecimento e uso do novo sistema que será implementado, assim como a importância do engajamento para a expansão da visibilidade intra e extra institucional do Repositório.

Referências

- CABALLERO RIVERO, A., SÁNCHEZ TARRAGÓ, N., SANTOS, R. N. M. D. (2019). Práticas de ciência aberta da comunidade acadêmica brasileira: estudo a partir da produção científica. *Transinformação*, 31. <https://doi.org/10.1590/2318-0889201931e190029>
- SARVO, D. O., FARIA, L. I. L., MATIAS, M. S. O., MATIAS, P., AMARAL, R. M. Ineficácia do povoamento automatizado de repositórios institucionais a partir da Web of Science. In: Encontro Nacional de Pesquisa em Ciência Da Informação (2016, 20 novembro) Salvador. <https://brapci.inf.br/index.php/res/v/191711>
- GODIN, B.; DORÉ, C. (2005). *Measuring the impacts of science: beyond the economic dimension*. INRS.
- VICENTE SAEZ., R.; MARTINEZ FUENTES, C. (2018). Open Science now: A systematic literature review for an integrated definition. *Journal of Business Research*, 88(C), 428-436.
- VIDOTTI, S. A. B. G., BASTOS, R. M., GRISOTO, A. P., ARAKAKI, F. A., FERREIRA, J. B. (2016, 17 de maio). Coleta automática para povoamento de repositórios digitais: conversão de registros utilizando XSLT. In: Encontro Nacional de Pesquisa em Ciência da Informação, Bahia. <http://hdl.handle.net/11449/144718>

Ampliação da ciência aberta na Universidade Federal de Minas Gerais: ações e perspectivas para ampla visibilidade dos resultados de pesquisa

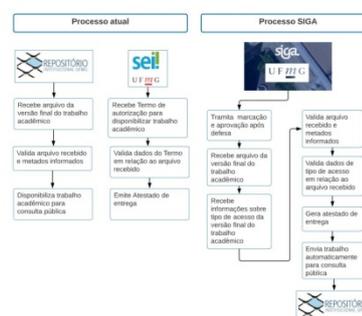
Izabel Antonina de Araújo
Universidade Federal de Minas Gerais, Brasil
antoninaizabel@gmail.com

Camila Mariana Aparecida da Silva
Universidade Federal de Minas Gerais, Brasil
camilamas.silva@gmail.com

Com o intuito de ampliar a ciência aberta na Universidade Federal de Minas Gerais, o trabalho descreve duas propostas como processos facilitadores para ampliar a visibilidade dos resultados de pesquisas e oferecer um gerenciamento mais efetivo da produção intelectual da instituição.

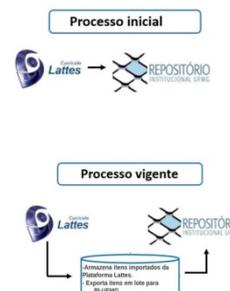
Primeira proposta: integração do Sistema de Gestão Acadêmica (SIGA) com o Repositório

Visa trazer soluções para problemas como morosidade na disponibilização dos trabalhos acadêmicos, segurança da informação e otimização dos fluxos de trabalho. Neste processo, todos os passos, desde a marcação da defesa até a disponibilização das teses, dissertações e monografias, serão realizadas em um único sistema. Haverá maior agilidade na disponibilização dos trabalhos para consulta pública e emissão de atestados de entrega, necessários à solicitação do diploma pelo egresso. Além de permitir a descentralização do fluxo de trabalho, hoje a cargo da equipe do Repositório, para as bibliotecas das unidades acadêmicas.



Segunda proposta: criação da base Importa Lattes para integração de itens importados ao Repositório

Busca solucionar o problema de carregamento do DSpace que não suportou o armazenamento de um número expressivo de itens em uma única fila para validação dos dados, especialmente quando ocorriam acessos simultâneos. Com a base Importa Lattes é possível armazenar dados referenciais dos 78 mil importados da Plataforma Lattes e cuja qualidade foi aprimorada após aplicação de algumas regras. O envio de itens para validação no Repositório passa a ser feito por lotes, diminuindo substancialmente a sobrecarga do Dspace. A base utiliza metadados principais de identificação dos itens como campo de busca da produção importada, além de permitir a associação de filtros que aumentam a precisão na recuperação.



O projeto SIGA encontra-se em desenvolvimento e foi considerado inovador pelos órgãos administrativos da Universidade, tendo sido solicitada prioridade em seu desenvolvimento em consideração ao alto nível de impacto em diferentes setores da instituição e por implicar em racionalização do fluxo vigente. Com o desenvolvimento do estudo, constatamos que para garantir celeridade na disponibilização dos resultados de pesquisa, evitar duplicidade de trabalhos, favorecer economia de recursos e o gerenciamento efetivo da produção intelectual, são necessários investimentos em infraestrutura informacional com unificação dos processos.



Captação de recursos para projetos desenvolvidos pelo Núcleo de Sistema Integrado de Bibliotecas – NSIB do INCA: Lições aprendidas

ROBSON DIAS MARTINS

Instituto Nacional de Câncer José Alencar Gomes da Silva

robson.martins@inca.gov.br

KÁTIA DE OLIVEIRA SIMÕES

Instituto Nacional de Câncer José Alencar Gomes da Silva

katia.simoese@inca.gov.br

CAMILA BELO TAVARES FERREIRA

Instituto Nacional de Câncer José Alencar Gomes da Silva

camila.ferreira@inca.gov.br

RESUMO

Eixo temático: Comunicação acadêmica e científica aberta.

As bibliotecas brasileiras sofrem com a escassez de recursos financeiros, humanos e físicos para o desenvolvimento de serviços e produtos de qualidade. A literatura brasileira em Biblioteconomia é escassa sobre a gestão e a captação de recursos para suas unidades. No cenário de escassez de estudos acerca do assunto, esse trabalho pretende apresentar o processo e as etapas para a captação de recursos em dois projetos realizados pela Núcleo de Sistema Integrado de Bibliotecas e a Biblioteca Virtual em Prevenção e Controle de Câncer, pertencentes à área de Ensino do Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA). Os projetos apresentados foram: Implementação do Repositório Institucional do INCA e; Gestão do processo de certificação da BVS Prevenção e

Controle de Câncer voltados para o Acesso Aberto à informação técnico-científica da área da saúde, mais especificamente, informações oncológicas de prevenção e controle do câncer. O processo de construção dos projetos envolverem parcerias estratégicas com outra área do Instituto, principalmente, na construção e desenvolvimento baseados no modelo PICTH, metodologia utilizada para o desenvolvimento e apresentação dos projetos. O trabalho pretende contribuir com a classe biblioteconômica apresentando novas ideias e ações para instrumentalizar os profissionais na elaboração de projetos.

PALAVRAS CHAVE

Captação de recursos; gerenciamento de projetos; políticas institucionais.



Captação de recursos para projetos desenvolvidos pelo Núcleo de Sistema Integrado de Bibliotecas – NSIB do INCA: Lições aprendidas

Robson Dias Martins

Instituto Nacional de Câncer José Alencar Gomes da Silva

robson.martins@inca.gov.br

Kátia de Oliveira Simões

Instituto Nacional de Câncer José Alencar Gomes da Silva

Katia.simoes@inca.gov.br

Camila Belo Ferreira

Instituto Nacional de Câncer José Alencar Gomes da Silva

camila.ferreira@inca.gov.br

INTRODUÇÃO

As bibliotecas brasileiras sofrem com a escassez de recursos financeiros, humanos e físicos para o desenvolvimento de serviços e produtos de qualidade. A literatura brasileira em Biblioteconomia é escassa sobre a gestão e a captação de recursos para suas unidades. No cenário de escassez de estudos acerca do assunto, esse trabalho pretende apresentar o processo e as etapas para a captação de recursos em dois projetos realizados pelo Núcleo de Sistema Integrado de Bibliotecas e a Biblioteca Virtual em Prevenção e Controle de Câncer, pertencentes à área de Ensino do Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA).

Os projetos apresentados foram: Implementação do Repositório Institucional do INCA e; Gestão do processo de certificação da BVS Prevenção e Controle de Câncer voltados para o Acesso Aberto à informação técnico-científica da área da saúde, mais especificamente, informações oncológicas de prevenção e controle do câncer.

Figura 1 – Interface BVS Prevenção e Controle de Câncer



Fonte: <https://controlecancer.bvs.br>

Figura 2 - Interface Repositório Institucional INCA



Fonte: Página interna/ acesso restrito

O processo de construção dos projetos envolverem parcerias estratégicas com outra área do Instituto, principalmente, na construção e desenvolvimento baseados no modelo PICTH, metodologia utilizada para o desenvolvimento e apresentação dos projetos.

O trabalho pretende contribuir com a classe biblioteconômica apresentando novas ideias e ações para instrumentalizar os profissionais na elaboração de projetos.

REFERÊNCIA

SOUZA, Andrea Batista de *et al.* Estratégia de captação de recursos para implementação de projetos: experiência da biblioteca pública do estado com a bersato produções. **XXVIII CBBB**, Vitória - ES. v. 28, p. 2-6, 2019. Disponível em: <https://portal.febab.org.br/anais/article/view/2302>. Acesso em: 22 out. 2021.



Modelo de arquivamento de páginas web para Portais de Periódicos: uma iniciativa experimental do Portal de Periódicos da UNICAMP

GILDENIR CAROLINO SANTOS

Universidade Estadual de Campinas

gilldenir@unicamp.br

<https://orcid.org/0000-0002-4375-6815>

JULIANO SIMIÃO CARDOZO

Universidade Estadual de Campinas

ppec@unicamp.br

RESUMO

Eixo temático: Infraestrutura tecnológica.

O avanço tecnológico provocou transformações no campo da informação, e com isso ocorreu uma grande explosão bibliográfica, que demandou tempo para organização e tratamento de documentos, além de modificar as formas de preservação e apropriação de conhecimentos para investir na preservação digital de páginas web. Nesse aspecto, o presente trabalho tem como objetivo apresentar uma iniciativa independente do projeto de preservação das páginas web do Portal de Periódicos Eletrônicos Científicos (PPEC) da Universidade Estadual de Campinas (UNICAMP), vinculado ao Sistema de Bibliotecas (SBU). O PPEC participa de redes de preservação, tais como Rede Carinia (Ibict) e a PKP PN, para buscar e garantir o acesso aberto aos seus periódicos. A partir deste contexto, para descrever a experiência de preservação de documentos, utilizou-se a metodologia experimental, com embasamento bibliográfico, para o arquivamento das páginas web do PPEC, por meio da ferramenta

Conifer, desenvolvida pela Rhizome, uma instituição sem fins lucrativos que apoia a arte e a cultura digital nata. O PPEC efetuou o cadastro no Conifer, e obteve 20 GB de espaço para hospedagem gratuita. O trabalho de captura é realizado por um bolsista sob a coordenação do bibliotecário responsável pelo PPEC. Atualmente estão preservadas as páginas dos 32 títulos de periódicos, além do site de gerenciamento de todo o PPEC, o software Open Journal System – OJS (PKP). A realização desta metodologia pelo Conifer, permite a captura das páginas para o arquivamento com a inclusão básica de metadados pré-definidos no sistema (data de captura, título da página, URL, session ID e tipo de browser). Como resultados, obtivemos até o momento, o arquivamento de 3 títulos de periódicos e das página principais do PPEC. O formato para download das páginas é o WARC, padrão de arquivo de container para armazenar conteúdo web em seu contexto original, mantido pelo International Internet Preservation Consortium (IIPC). Considera-se que este trabalho demandará tempo para sua finalização, mas já é perceptível sua relevância para a preservação e o arquivamento das páginas web do PPEC.

PALAVRAS CHAVE

Preservação web; arquivamento de páginas web; Conifer (ferramenta de arquivamento); portais de periódicos.

Web preservation; web archiving; Connifer (archiving tool); journal portals.



Modelo de arquivamento de páginas web para Portais de Periódicos: uma iniciativa experimental do Portal de Periódicos da UNICAMP

SANTOS, Gildenir Carolino. Universidade Estadual de Campinas, SBU/PPEC, Brasil
CARDOZO, Juliano. Universidade Estadual de Campinas, SBU/PPEC, Brasil

Palavras-chaves

Preservação web. Arquivamento de páginas web. Conifer (Ferramenta de arquivamento). Portais de periódicos.

Eixo temático

Infraestrutura tecnológica

Resumo

Introdução: O avanço tecnológico provocou transformações no campo da informação, e com isso ocorreu uma grande explosão bibliográfica, que demandou tempo para organização e tratamento de documentos, além de modificar as formas de preservação e apropriação de conhecimentos para investir na preservação digital de páginas web. **Objetivo:** Nesse aspecto, o presente trabalho tem como objetivo apresentar uma iniciativa independente do projeto de preservação das páginas web do Portal de Periódicos Eletrônicos Científicos (PPEC) da Universidade Estadual de Campinas (UNICAMP), vinculado ao Sistema de Bibliotecas (SBU). O PPEC participa de redes de preservação, tais como Rede Carinia (Ibict) e a PKP PN, para buscar e garantir o acesso aberto aos seus periódicos. **Metodologia:** A partir deste contexto, para descrever a experiência de preservação de documentos, utilizou-se a metodologia experimental, com embasamento bibliográfico, para o arquivamento das páginas web do PPEC, por meio da ferramenta Conifer,

desenvolvida pela Rhizome, uma instituição sem fins lucrativos que apoia a arte e a cultura digital nata. O PPEC efetuou o cadastro no Conifer, e obteve 20 gb de espaço para hospedagem gratuita. O trabalho de captura é realizado por um bolsista sob a coordenação do bibliotecário responsável pelo PPEC.

Atualmente estão preservadas as páginas dos 32 títulos de periódicos, além do site de gerenciamento de todo o PPEC, o software Open Journal System – OJS (PKP). A realização desta metodologia pelo Conifer, permite a captura das páginas para o arquivamento com a inclusão básica de metadados pré-definidos no sistema (data de captura, título da página, URL, session ID e tipo de browser). **Resultados:** Como resultados, obtivemos até o momento, o arquivamento de 3 títulos de periódicos e das páginas principais do PPEC. O formato para download das páginas é o WARC, padrão de arquivo de container para armazenar conteúdo web em seu contexto original, mantido pelo International Internet Preservation Consortium (IIPC). **Conclusão:** Considera-se que este trabalho demandará tempo para sua finalização, mas já é perceptível sua relevância para a preservação e o arquivamento das páginas web do PPEC.

Licença



Experiencia en la gestión de la productividad académica a través de colecciones dinámicas de LENS.ORG en la Universidad de Guadalajara, un estudio transversal

NETZAHUALPILLI DELGADO FIGUEROA

Universidad de Guadalajara

netzahualpilli.delgado@udgvirtual.udg.mx

GUADALUPE JEANETTE GONZÁLEZ DÍAZ

Universidad de Guadalajara

jeanette.gonzalez@udgvirtual.udg.mx

ESMERALDA ALCARAZ SÁNCHEZ

Universidad de Guadalajara

biol.esmeralda@gmail.com

MARTÍN ADALBERTO TENA ESPINOZA DE LOS MONTEROS

Universidad de Guadalajara

mtenaespinoza@udgvirtual.udg.mx

RESUMEN

Eje temático: Infraestructura tecnológica.

La gestión de la investigación es una necesidad de las instituciones educativas universitarias, muchas de ellas han realizado esfuerzos por mantener repositorios institucionales, aunque la mayoría de ellos carecen de automatización, interoperabilidad, metadatos y visualizaciones. El objetivo de este proyecto es presentar la experiencia del uso de colecciones dinámicas creadas con la plataforma LENS.ORG y contrastar

las frecuencias de obras reportadas en los perfiles de ORCID de los investigadores.

Materiales y metodología: En este estudio de diseño transversal se explora la recuperación de la productividad académica del campus universitario usando las colecciones dinámicas generadas desde LENS.ORG, usando como Query «Author Display Name» y enlazadores booleanos OR, dado que existen limitaciones importantes en el uso de identificadores persistentes únicos como ORCID por parte de los investigadores de la institución, fue utilizado el padrón de investigadores con nombramiento del Sistema Nacional de Investigadores (SNI-CONACYT) del Centro Universitario de Ciencias Biológicas y Agropecuarias con un universo de 119 investigadores, realizándose la solicitud manual de obras reportadas en ORCID y comparándolas con los resultados de LENS.ORG en las búsquedas dinámicas se reportan frecuencias y proporciones.

Resultados: A través de la creación de colecciones dinámicas fue posible identificar un total de 968 obras, representando un acercamiento efectivo para instituciones con limitaciones del uso de identificadores únicos como ORCID, en este caso fue solo del 63,8 % (n=76).

PALABRAS CLAVE

CRIS; información de investigación; estudio transversal.

CRIS; research information; cross-sectional study.

Bibliografía

- JEFFERSON, O. A., KOELLHOFER, D., WARREN, B., and JEFFERSON, R. (2019, November 25). The Lens MetaRecord and LensID: An open identifier system for aggregated metadata and versioning of knowledge artefacts. <https://doi.org/10.31229/osf.io/t56yh>
- JEFFERSON, O. A., ADAM JAFFE, DOUG ASHTON, BEN WARREN, DENIZ KOELLHOFER, UWE DULLECK, AARON BALLAGH, JOHN MOE, MICHAEL DiCUCCIO, KARL WARD, GEOFF BILDER, KEVIN DOLBY, AND RICHARD JEFFERSON. (2018). Mapping the global influence of published research on industry and innovation. *Nature Biotechnology* 36, 31-39. <https://www.lens.org/036-630-785-708-925> <https://doi.org/10.1038/nbt.4049>

JEFFERSON, R. (2017). Comment: Turning science into social outcomes. *Nature* 548, S8.
<https://www.lens.org/020-882-124-317-089> <http://dx.doi.org/10.1038/548s8a>

Experiencia en la gestión de la productividad académica a través de colecciones dinámicas de LENS.ORG en la Universidad de Guadalajara, un estudio transversal.

Delgado-Figueroa Netzahualpilli¹, netzahualpilli.delgado@udgvirtual.udg.mx; González-Díaz Guadalupe Jeanette¹, jeanette.gonzalez@udgvirtual.udg.mx; Alcaraz-Sánchez Esmeralda¹, biol.esmeralda@gmail.com; Tena-Espinoza-De-los-Monteros Martín Adalberto¹, mtenaespinoza@udgvirtual.udg.mx.
¹Universidad de Guadalajara, México.



Introducción:

Las instituciones buscan implementar mecanismos que les permitan medir, organizar y calificar la calidad de la productividad científica generada, convirtiendo a la gestión de la investigación en una necesidad, gran parte de los procesos realizados en muchas instituciones carecen de automatización, interoperabilidad, metadatos y visualizaciones. Organizaciones como LENS.ORG poseen herramientas que facilitan la organización, procesamiento, adición de metadatos de las obras académicas generadas en las instituciones. Una de estas herramientas es la generación de colecciones dinámicas.

Resultados:

A través de la creación de colecciones dinámicas fue posible identificar un total de 968 obras académicas únicas entre los autores, representando un acercamiento efectivo para instituciones con limitaciones en la frecuencia del uso de identificadores únicos como ORCID, en este caso fue solo del 63.8% (n=76) la cantidad de autores identificados con ORCID del padrón de investigadores, los resultados de una búsqueda manual por ORCID fue de 1,209 obras académicas, teniendo como limitante que desde la búsqueda en ORCID es frecuente las duplicaciones de documentos entre autores (19.9%).

Materiales y Metodología:

En este estudio de diseño transversal se explora la recuperación de la productividad académica del Centro Universitario de Ciencias Biológicas y Agropecuarias usando las colecciones dinámicas generadas desde el sistema LENS.ORG, usando como Query para las búsquedas "Author Display Name" y operadores booleanos "OR", esta estrategia fue generada al observar que existen limitaciones importantes en el uso de identificadores persistentes únicos como ORCID por parte de los investigadores de la institución, fue utilizado el padrón de investigadores con nombramiento del Sistema Nacional de Investigadores (SNI-CONACYT) del Centro Universitario de Ciencias Biológicas y Agropecuarias con un universo de 119 investigadores, realizándose la solicitud manual de obras reportadas en ORCID en cada uno de los perfiles de autor y comparándolas con los resultados de LENS.ORG en las búsquedas dinámicas, fueron reportadas frecuencias y proporciones de las obras académicas obtenidas.

Conclusiones:

Al utilizar LENS.ORG representa para las instituciones un acercamiento para identificar obras únicas y mantener repositorios con búsquedas automatizadas, agilizando el proceso de mantener repositorios institucionales, el proceso de implementación de estas herramientas puede ser un elemento clave para la creación de sistemas CRIS en las instituciones.



Bibliografía:

- Jefferson, O. A., Koellhofer, D., Warren, B., and Jefferson, R. (2019, November 25). The Lens MetaRecord and LensID: An open identifier system for aggregated metadata and versioning of knowledge artefacts. (<https://doi.org/10.31229/osf.io/t56yh>)
- Jefferson, O.A., Adam Jaffe, Doug Ashton, Ben Warren, Deniz Koellhofer, Uwe Dulleck, Aaron Ballagh, John Moe, Michael DiCuccio, Karl Ward, Geoff Bilder, Kevin Dolby, and Richard Jefferson (2018) Mapping the global influence of published research on industry and innovation. Nature Biotechnology 36, 31-39. (<https://www.lens.org/036-630-785-708-925>, <https://doi.org/10.1038/nbt.4049>)
- Jefferson, R. (2017). Comment: Turning science into social outcomes. Nature 548, S8. (<https://www.lens.org/020-882-124-317-089>, <http://dx.doi.org/10.1038/548s8a>)



Repositorio Centroamericano de Patrimonio Cultural del IIARTE (UCR) como espacio para divulgar colecciones de obras utilizadas en investigación

SOFÍA VINDAS SOLANO

Instituto de Investigaciones en Arte (IIARTE), Universidad de Costa Rica

sofia.vindas@ucr.ac.cr

MEILYN GARRO ACÓN

Vicerrectoría de Investigación, Universidad de Costa Rica

meilyn.garro@ucr.ac.cr

RESUMEN

El proyecto nació en 2014 como un espacio para recopilar y divulgar bases de obras de acceso libre sobre patrimonio cultural centroamericano. En un principio, el proyecto fue pensado para la recopilación de patrimonio documental, sin embargo, al día de hoy se albergan bases de artes visuales y arte sonoro.

Metodología:

- Identificación de proyectos de investigación que trabajan con bases de obras de patrimonio cultural que necesitan un espacio para almacenar y organizar las obras recopiladas y que pueden compartir su base de datos al final de la investigación.
- Acompañamiento para personas investigadoras en la creación de colecciones digitales sobre patrimonio cultural. Incluye asesoría en procesos de digitalización, gestión de derechos de autor, catalogación y normalización de bases de datos para su divulgación e interoperabilidad.
- Responsabilidad compartida con investigadores y unidades académicas para el mantenimiento y actualización de las colecciones.

- Organización y mantenimiento de un espacio web para divulgar las colecciones. Actualmente se trabaja con Dspace repositorio.iiarte.ucr.ac.cr

Resultados:

- Se cuenta con más de 6000 items de artes visuales, arte sonoro y patrimonio documental, provenientes de Costa Rica, Guatemala, El Salvador y Panamá.
- Se mantiene una convocatoria abierta para nuevas colecciones.

PALABRAS CLAVE

Acceso abierto; patrimonio cultural; red colaborativa.



Comunicación académica y científica en abierto

Repositorio Centroamericano de Patrimonio Cultural, del IIArte (UCR) como espacio para divulgar colecciones de obras utilizadas en investigación

Sofía Vindas Solano
Universidad de Costa Rica, Instituto de Investigaciones en Arte
Meilyn Garro Acón
Universidad de Costa Rica, Vicerrectoría de Investigación

Palabras clave: acceso abierto, patrimonio cultural, red colaborativa

Objetivo:

El proyecto nació en la Universidad de Costa Rica (UCR), en 2014, como un espacio para recopilar y divulgar bases de obras de acceso libre sobre patrimonio cultural centroamericano. En un principio, el proyecto fue pensado para la recopilación de patrimonio documental, sin embargo, al día de hoy se albergan bases de artes visuales y arte sonoro.

Metodología:

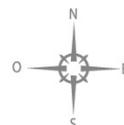
- Identificación de proyectos de investigación que trabajan con bases de obras de patrimonio cultural que necesitan un espacio para almacenar y organizar las obras recopiladas y que pueden compartir su base de datos al final de la investigación.
- Acompañamiento para personas investigadoras en la creación de colecciones digitales sobre patrimonio cultural. Incluye asesoría en procesos de digitalización, gestión de derechos de autor, catalogación y normalización de bases de datos para su divulgación e interoperabilidad.
- Responsabilidad compartida con investigadores y unidades académicas para el mantenimiento y actualización de las colecciones.
- Organización y mantenimiento de un espacio web para divulgar las colecciones. Actualmente se trabaja con Dspace repositorio.iiarte.ucr.ac.cr

Resultados:

- Se cuenta con más de 6000 ítems de artes visuales, arte sonoro y patrimonio documental, provenientes de Costa Rica, Guatemala, El Salvador y Panamá (datos de julio de 2021).
- Según Google Analytics 12/07/21 (<https://repositorio.iiarte.ucr.ac.cr/page/estadisticas>):
 - Se recibe un promedio de 300 mil visitas mensuales.
 - Los 5 de países que registran más visitas son EEUU, India, Canadá, Reino Unido, Japón.
 - El 56,5% de los usuarios son mujeres, mientras que el 43,5% son hombres.
 - Se registra un promedio de 45 mil usuarios mensuales, de los cuales la mayoría se encuentra entre las edades de 18 a 36 años.
- El proyecto forma parte de una red de sitios (entre bibliotecas, repositorios y universidades) centroamericanos con los que se negocian estándares y se establecen redes de cooperación.
- No se ha identificado otro proyecto similar en la región, que reúna diversidad de investigaciones sobre patrimonio cultural.



Info de contacto:
patrimoniocultural@ucr.ac.cr
Instagram: @elrepoiarte
Facebook: <https://www.facebook.com/>



Todos los derechos reservados Universidad de Costa Rica
Atribución - No Comercial - Compartir bajo la Misma Licencia Costa Rica 3.0

Otras actividades

Consortio Iberoamericano para Educación en Ciencia y Tecnología (ISTEC). Nuevos proyectos

MARISA RAQUEL DE GIUSTI

Consortio Iberoamericano para Educación en Ciencia y Tecnología

marisa.degiusti@sedici.unlp.edu.ar

RESUMEN

Esta presentación estuvo dedicada a poner en contexto a ISTEC, su misión, visión, alcances y socios, para luego enfocarse en los nuevos proyectos. Para iniciar la charla se centró en la Iniciativa LibLink del ISTEC y en particular en su plan de acciones para 2021 y 2022. La iniciativa reconoce como objetivos y acciones permanentes aquellas vinculadas al intercambio de recursos bibliográficos, que aunque tiende a disminuir en cantidad, se enfoca en materiales muy difíciles y por lo tanto exige de personas (referencistas) altamente capacitadas. Para la gestión bibliográfica y compartición de contenidos, ISTEC, a través del PREBI-SEDICI de la UNLP, mantiene el software Celsius y atiende la formación de los socios en el uso de este desarrollo. Para este período agrega, a lo precedente, la difusión, formación y capacitación de recursos humanos en América Latina en Ciencia Abierta, así como la difusión de un modelo incremental y sostenible. Durante 2022 la iniciativa pretende comenzar un ciclo de capacitación abierta que se llamará «Hilos de Conocimiento» destinado a compartir con los socios experiencias sobre distintos tópicos de la ciencia abierta: fundamentos y componentes en Ciencia Abierta, estrategias, gestión de datos de investigación, ciencia colaborativa y evaluación abierta, entre otros temas. Otro proyecto sobre el cual se conversó en esta presentación es la puesta a punto de un procedimiento de auditoría y evaluación de repositorios institucionales de acceso abierto. La definición de esta propuesta avanzará en el transcurso de 2022 de acuerdo al interés

que manifiesten los socios. El lanzamiento de un servicio de mejora de calidad de revistas de las instituciones miembro, adecuación al acceso abierto e indexación que está listo para su comienzo de acuerdo a las necesidades manifestadas por los socios.

PALABRAS CLAVE

Educación; tecnología; capacitación; bibliografía; ciencia abierta.

DiNaRe: El nuevo servicio de REMERI para la visibilidad del Acceso Abierto en México

ROSALINA VÁZQUEZ TAPIA

Universidad Autónoma de San Luis Potosí

alinamte@gmail.com

ANTONIO FELIPE RAZO RODRÍGUEZ

Universidad de las Américas Puebla

antrazo@cudi.edu.mx

RESUMEN

La Red Mexicana de Repositorios Institucionales (REMERI) fue fundada en el 2012 con el propósito fundamental de promover el desarrollo de los repositorios institucionales y temáticos de acceso abierto en México y posibilitar su consulta a través de una interfaz común basada en búsquedas federadas y estándares internacionales de interoperabilidad. Adicionalmente, desde esa fecha, REMERI ha sido el nodo técnico de México en la Red Federada de Repositorios Institucionales de Publicaciones Científicas de América Latina (LA Referencia). Con el objetivo de contar con un registro estandarizado de los repositorios digitales de México y brindar mayores servicios de información a los responsables o administradores de los repositorios integrados a REMERI, en 2021 se emprende el desarrollo de un nuevo proyecto denominado Directorio Nacional de Repositorios Digitales (DiNaRe). Los servicios proyectados para este Directorio son los siguientes: Formulario de registro de seis tipos de repositorios digitales (institucionales, temáticos, datos abiertos, recursos educativos, patrimonio documental y agregadores de revistas); Búsqueda y navegación por diferentes criterios; Generación y visualización de estadísticas; Modulo de administración por cuentas de acceso para responsables de repositorios, a través del cual podrán

actualizar sus datos, consultar materiales de ayuda, generar estadísticas y reportes sobre cosechas, transformaciones e indexaciones a sus repositorios. A la fecha (octubre 2021), a través del portal de REMERI están disponibles alrededor de 734.000 documentos a texto completo de 133 repositorios digitales correspondientes a 78 instituciones mexicanas de educación superior e investigación. Está previsto que el servicio de DiNaRe entre en operación en junio de 2022.

PALABRAS CLAVE

Repositorios digitales; acceso abierto; directorio de repositorios; red federada de repositorios.

Autores

ALBUQUERQUE, PABLO CÉSAR DE Licenciado en Sistemas en la Facultad de Informática de la Universidad Nacional de La Plata (UNLP). Actualmente, Becario doctoral CIC.

ALCARAZ SÁNCHEZ, ESMERALDA Maestra en Ciencias de la Salud Ambiental por la Universidad de Guadalajara (UDG), interés en la gestión del conocimiento y profesora universitaria con experiencia en investigación medio ambiental sobre los efectos de los contaminantes del aire y enfermedades alérgicas.

ALMEIDA FREYRE, ÉDER DE Mestre em Ciência da Informação pela Universidade Federal Fluminense (UFF) (2011). Especialista em Informação e Informática em Saúde pela Escola Nacional de Saúde Pública Sergio Arouca (2004). Graduado em Biblioteconomia e Documentação pela Universidade Federal Fluminense (UFF) (1998). Foi bibliotecário de sistemas na Biblioteca de Ciências Biomédicas do Instituto de Comunicação e Informação Científica e Tecnológica em Saúde (2004-2006). Foi responsável pelo setor de Processamento Técnico de Monografias e Multimeios da Biblioteca de Ciências Biomédicas do Instituto de Comunicação e Informação Científica e Tecnológica em Saúde (2007-2011). Foi Gestor de Comunidades e Coleções do Repositório Institucional da Fiocruz - ARCA (2011-2014). Técnico em Saúde Pública da Fundação Oswaldo Cruz (RJ) atualmente desenvolve suas atividades profissionais no Centro de Tecnologia de Informação e Comunicação, Seção de Informação - Arca - Repositório Institucional da Fiocruz. Participante ativo da Sudeste/RIAA de Repositórios Institucionais (IBICT) desde o ano de 2017. Um dos organizadores das Reuniões da Rede Sudeste de Repositórios Institucionais desde a edição 5 em 2018. Bibliotecário registrado no CRB-7 4913.

AMARO, BIANCA Doctora en Lingüística Aplicada por la Universidad Pompeu Fabra (España) y Licenciada en Derecho y Letras y más recientemente en Bibliotecología por la Universidade de Brasília (Brasil). Trabaja en los siguientes temas: Comunicación científica, Copyright, Acceso abierto a la información científica y Ciencia abierta. Coordinadora del Programa Brasileño de Acceso Abierto en el Instituto Brasileño de Información en Ciencia y Tecnología. Coordina los siguientes sistemas y proyectos: Portal de Publicaciones Científicas de Acceso Abierto (OasisBr); Biblioteca Digital Brasileña de Tesis y Disertaciones (BDTD); Directorio de Políticas de Acceso Abierto de Revistas Científicas Brasileñas (Diadorim); y los proyectos Repositorios Digitales de Acceso

	<p>Abierto y Revistas Científicas Brasileñas de Acceso Abierto. Co-organizadora de las Conferencias Luso-Brasileñas de Acceso Abierto (ConfOA). Miembro del Consejo Ejecutivo de la Confederación de Repositorios de Acceso Abierto (COAR). Ganadora del premio International Electronic Publishing Trust for Development (EPT 2015). Presidente de la Red de Acceso Abierto a Repositorios Científicos - LA Referencia.</p>
AQUINO REMIGIO, SILVANA	<p>Licenciada en Bibliotecología y Ciencias de la Información por la Universidad Nacional Mayor de San Marcos (UNMSM) y magíster en Sistemas de Información Digital por la Universidad de Salamanca (USAL). Asimismo, fue becada para realizar pasantías profesionales y curso referido a bibliotecas digitales en España. Su experiencia profesional y especialización se ha centrado en los últimos años en los ámbitos de la gestión de la información especializada en CTI y la gestión de repositorios. Fue responsable del Repositorio Digital de Recursos Hídricos de la Autoridad Nacional del Agua del Perú y actualmente se desempeña como Analista en Gestión de Repositorios Digitales en el Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica, teniendo entre sus funciones la gestión del Repositorio Institucional y el seguimiento y monitoreo de los repositorios adheridos al Repositorio Nacional ALICIA.</p>
ARAÚJO, IZABEL ANTONINA DE	<p>Possui graduação em Biblioteconomia pela Universidade Federal de Minas Gerais (UFMG); mestre em Ambiente Construído e Patrimônio Sustentável pela Escola de Arquitetura (UFMG). Doutora em Ciência da Informação pela Escola de Ciência da Informação (UFMG). Coordenadora do Repositório Institucional da UFMG.</p>
ATKINSON, JOHN	<p>Ingeniero Civil y Máster en Ciencias de la Computación por la Universidad Técnica Federico Santa María (UTFSM). En 2004 obtuvo su Doctorado en Inteligencia Artificial por la Universidad de Edimburgo, Escocia. En la Facultad de Ingeniería y Ciencias se desempeña como profesor titular de Ingeniería Civil Informática donde imparte clases en pre y postgrado. Además, es director de Diplomado en Inteligencia Artificial que imparte la facultad desde 2017. Sus áreas de investigación abordan temáticas de las Ciencias de la Computación como el proceso natural del lenguaje, la extracción de textos, la inteligencia Artificial, entre otras. En 2010 recibió el ACM Senior Member Award por su contribución internacional a las Ciencias de la Computación. En los últimos 25 años ha estado involucrado en la enseñanza, investigación y consultoría, donde ha liderado proyectos científicos y tecnológicos. Fue profesor del Departamento de Ciencias de la Computación en la Universidad de Concepción, donde también fundó y dirigió el Laboratorio de Inteligencia Artificial. Previamente fue docente en la Universidad Técnica Federico Santa María (UTFSM).</p>
AZRILEVICH, PAOLA ALEJANDRA	<p>Especialista en gestión de la información y el conocimiento. Entre 1998-2005 trabajó en el CAICYT-CONICET, como asistente del Centro Argentino del ISSN, donde participó de proyectos como LATINDEX y SciELO. Entre el 2005-2008 fue directora de la Biblioteca de AIDIS ARGENTINA, coordinadora</p>

	<p>nacional de la Red Panamericana de Información en Salud Ambiental y su Biblioteca Virtual de Desarrollo Sostenible y Salud Ambiental (REPIDISCA/BVSDE ARGENTINA) e integrante del Comité Consultivo Nacional de la Biblioteca Virtual en Salud Argentina. Fue además, consultora de la OPS/OMS para el desarrollo de la Biblioteca Virtual de Municipios y Comunidades Saludables. En la actualidad es responsable de la Secretaría Ejecutiva del Sistema Nacional de Repositorios Digitales y de la Biblioteca Electrónica de Ciencia y Tecnología y del Ministerio de Ciencia, Tecnología e Innovación Productiva.</p>
BARRETO MALHEIRO PEREIRA, CATARINA	<p>Graduanda em Biblioteconomia e Gestão de Unidades de Informação pela Universidade Federal do Rio de Janeiro (UFRJ), cursando o oitavo (8º) período. Atualmente é estagiária da Fundação Oswaldo Cruz, lotada no Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, atuando com a equipe executiva do Repositório Institucional Arca. Possui experiência em curadoria de dados e repositório institucional.</p>
BECERRIL GARCÍA, ARIANNA	<p>Profesora-investigadora de tiempo completo en la Universidad Autónoma del Estado de México (UAEM). Miembro del Sistema Nacional de Investigadores de México. Cuenta con estudios de posgrado, doctora y maestra en ciencias de la computación por parte del Tecnológico de Monterrey, México. Ingeniera en Computación (UAEM). Es miembro del equipo fundador de la Red de Revistas Científicas de América Latina y el Caribe, España y Portugal (redalyc.org), y su actual Directora Ejecutiva. Es fundadora y directora de AmeliCA Conocimiento Abierto S. C. Además, es cofundadora de la Red Mexicana de Repositorios Institucionales y forma parte del comité; directivo de Invest In Open Infrastructure, es miembro del consejo de The Global Sustainability Coalition for Open Science Services (SCOSS) así; como del consejo de Directory of Open Access Journals (DOAJ).</p>
BELCHIOR RODRIGUES, RAPHAEL	<p>Graduando em Biblioteconomia e documentação pela Universidade Federal Fluminense (UFF). Atualmente é Bolsista da Fundação Oswaldo Cruz, atuando com a equipe executiva do repositório institucional Arca.</p>
BELO TAVARES FERREIRA, CAMILA	<p>Mestre em Ciência da Informação pelo Programa de Pós-Graduação em Ciência da Informação do IBICT/UFRJ (2012). Graduou-se em Biblioteconomia e Documentação na Universidade Federal Fluminense (UFF) (2007). É Tecnologista Pleno (Especialidade: Divulgação Científica em Saúde) do Instituto Nacional de Câncer (INCA). Atua no Sistema Integrado de Bibliotecas e no Serviço de Educação e Informação Técnico Científica do INCA. É supervisora técnica da Biblioteca Virtual Prevenção e Controle do Câncer. Possui experiência na área de Saúde Baseada em Evidências e Avaliação de Tecnologias em Saúde. Desenvolve atividades ligadas às bases de dados científicas, gestão da informação em saúde, diretrizes clínicas baseadas em evidências e revisão sistemática.</p>

BLANCO CASTILLO, HUMBERTO	Jefe de Innovación y Proyectos del CRAI en la Universidad del Rosario (UR), Colombia. Ingeniero de sistemas, especialista en gerencia de proyectos TIC. Experto en el desarrollo de proyectos enfocados a la implementación, visibilidad e interoperabilidad de repositorios institucionales, así como el desarrollo de soluciones basadas en software libre para la gestión de bibliotecas. Actualmente lidera las estrategias para promover la visibilidad de la producción institucional en acceso abierto, la generación iniciativas y gestión proyectos de base tecnológica que apoyan a los procesos de innovación del CRAI.
BOCANEGRA ESQUEDA, TOMÁS	Doctor en Bibliotecología y Estudios de la Información (UNAM). Sus áreas de interés son la metodología de investigación, en particular los estudios comparados, servicios al usuario y alfabetización en información. Actualmente se desempeña como bibliógrafo asignado al Centro de Estudios Económicos en la BDCV de El Colegio de México. También es vicepresidente del Colegio Nacional de Bibliotecarios A. C.
CADENAS MARÍN, LUIS ELIÉCER	De nacionalidad venezolana y reconocida experiencia en el mundo de las redes académicas y la escena TIC latinoamericana, es el Director Ejecutivo de RedCLARA desde el 1 de agosto de 2017. Ingeniero en computación con estudios de postgrado en economía, telemática, redes y ciencias de la computación, con casi treinta años de experiencia en el sector de tecnologías de la información y telecomunicaciones en el ámbito privado y público, en su carrera profesional ha buscado ser un agente de cambio, crecimiento y desarrollo de las personas y organizaciones mediante las tecnologías de información y comunicación. En esta línea, fue fundamental su rol como cofundador, presidente y miembro del consejo directivo de la Red Académica Nacional de Venezuela (REACCIUN), en la que tuvo participación central entre los años 1996 y 2000. Como presidente de REACCIUN, logró desarrollar la primera red física nacional (financiada con fondos del Banco Interamericano de Desarrollo), conectando a 17 universidades públicas, incrementando la disponibilidad de ancho de banda para la comunidad científico-académica y haciendo de éste un servicio estable y fiable. Fue presidente de la Asociación Latinoamericana de Red de Investigación (Enredo 1998-1999), miembro del directorio del Centro Nacional de Supercómputo de Venezuela (Cecalcula, 1997-1999), y del directorio del Parque Tecnológico de Sartenejas (1997-1999). Al moverse al ámbito de la empresa privada, los primeros años se mantuvo relacionado con el espectro académico al liderar la estrategia para el Programa Académico de Redes de Cisco (2000-2006), luego, motivado por sumar nuevos conocimientos, se movió dentro de Cisco hacia el ámbito comercial ocupando el cargo de Gerente de Cuentas para el sector público de Venezuela (2006-2009), luego para los proveedores de servicio (2009-2013), hasta que en 2014 se mudó junto a su familia a Guatemala para liderar la estrategia comercial de la empresa con América Móvil en América Central. Hasta asumir como Director Ejecutivo de RedCLARA en 2017, ocupaba la posición de Gerente de Cuenta para Claro América Central en Cisco.

CANTO LEAL, LAURA REGINA DO	Graduada em Biblioteconomia pela Universidade Federal do Rio Grande do Sul (UFRGS) (2020) e em Licenciatura em História pelo Centro Universitário Metodista (2008), com Pós-Graduação em História do Brasil Contemporâneo, pela Faculdade Porto-Alegrense (2010). Mestranda do Programa de Pós-Graduação em Ciência da Informação (PPGCIN-UFRGS). Área de interesse: Ciência da Informação, Ciência Cidadã, Comunicação Científica e Produção Científica.
CAREGNATO, SÔNIA ELISA	Possui graduação em Biblioteconomia pela Universidade Federal do Rio Grande do Sul (UFRGS), mestrado em Information Management pela University of Sheffield e doutorado em Information Studies também pela University of Sheffield. É professora titular da Faculdade de Biblioteconomia e Comunicação (FABICO). É docente permanente no Programa de Pós-Graduação em Comunicação (PPGCOM) e no Programa de Pós-Graduação em Informação (PPGCIN), ambos da UFRGS. Foi presidente do Comitê Gestor do LUME - Repositório Digital da Universidade Federal do Rio Grande do Sul e diretora do CEDAP - Centro de Documentação e Acervo Digital da Pesquisa, órgão auxiliar da FABICO/UFRGS. Coordenou o GT7 - Produção e Comunicação de Informação em CT&I, da Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação - ANCIB, do qual faz parte. Atualmente, é coordenadora do Grupo de Pesquisa em Comunicação Científica da UFRGS. Tem experiência nas áreas de Ciência da Informação e Comunicação, atuando principalmente nos seguintes temas: comunicação científica, bibliometria, periódicos científicos, produção científica, repositórios digitais e acesso aberto a dados de pesquisa.
CASTELLANOS BOLAÑOS, MARÍA ENRIQUETA	Maestra en Gestión de Tecnología de Información por la Universidad Anáhuac Mayab, México. Es profesora de tiempo completo en la Facultad de Matemáticas de la Universidad Autónoma de Yucatán, México. Su trabajo de investigación se centra en temas relacionados con la gestión del conocimiento y la ingeniería web.
CID CARMONA, VÍCTOR	Doctor en Bibliotecología y Estudios de la Información por la Universidad Nacional Autónoma de México (UNAM). Sus temas de interés se refieren particularmente a la historia del libro y la imprenta, sobre los que ha publicado: <i>El patrimonio bibliográfico angelopolitano en Chile y España</i> , <i>Repertorio de impresos mexicanos en la Biblioteca Nacional de España, siglos XVI-XVII</i> y <i>Epítome bibliográfico de impresos médicos mexicanos, siglos XVI-XVIII</i> , entre otros. Actualmente se desempeña como bibliógrafo asignado al Centro de Estudios Históricos en la Biblioteca Daniel Cosío Villegas de El Colegio de México.
COLMENARES MORENO, WILSON ENRIQUE	Magíster en Estudios Editoriales del Instituto Caro y Cuervo y profesional en Sistemas de Información, Bibliotecología y Archivística de La Universidad de La Salle. Ha sido editor técnico de libros y revistas científicas. Actualmente es investigador del Observatorio Editorial Colombiano del Instituto Caro y se especializa en la formulación, implementación y evaluación de proyectos editoriales,

bibliotecarios y de sistemas de organización del conocimiento para la apropiación social de la información cultural, científica y gubernamental a través de tecnologías de la información. Es autor del artículo de investigación «Aproximación al concepto de observatorio editorial: recomendaciones para su creación», 2021; autor del artículo «Importancia de la apropiación social y el acceso al conocimiento abierto especializado en Ciencias Agrarias», 2017; coautor libro *Divulgación y difusión del conocimiento: las revistas científicas*, 2016 y del capítulo de libro *Las publicaciones oficiales de la administración pública de Bogotá, Colombia*, 2015.

DE GIUSTI, MARISA RAQUEL

Doctora en Ciencias Informáticas, Ingeniera en Telecomunicaciones y Profesora en Letras (UNLP), asimismo se desempeña como Investigador Principal de la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC), donde además oficia como Coordinadora de la Comisión Asesora de Ingeniería, Arquitectura y otras Tecnologías. Es Profesora de Posgrado en la Facultad de Informática (UNLP). A fines de 1996 presentó a las autoridades de la Secretaría de Ciencia y Técnica de la UNLP el Proyecto de Enlace de Bibliotecas (PREBI), luego creado en mayo de 1997, desempeñándose desde entonces como su directora. De modo similar en el año 2001 presentó el Proyecto Servicio de Difusión de la Creación Intelectual (SEDICI), el cual se incorporó a PREBI y dirige desde su creación. SEDICI es el Repositorio Institucional de la Universidad Nacional de La Plata, creado para albergar, preservar y dar visibilidad a las producciones de sus Unidades Académicas y colegios. PREBI-SEDICI realiza numerosas actividades de I+D+I que forman parte del Plan Estratégico de la UNLP desde 2004. Directora del Centro de Servicios en Gestión de Información (CESGI), centro propio de la CIC dedicado a la gestión de la información científica, que entre otros programas y proyectos, gestiona el repositorio institucional CIC Digital. Desde 1990 trabaja con el Consorcio Iberoamericano para Educación en Ciencia y Tecnología (ISTEC), y en 2019 asume la presidencia por voto unánime del Consejo de Directores. Actualmente, es miembro del Directorio y Directora de la Iniciativa Library Linkage (LibLink) de ISTEC. Por su experiencia en la temática de Repositorios Digitales, integra el Comité de Expertos del Sistema Nacional de Repositorios Digitales (SNRD) del Ministerio de Ciencia, Tecnología e Innovación Productiva de la República Argentina. Cuenta con más de cien publicaciones internacionales dedicadas a las áreas de la gestión de la información, preservación digital, rankings y visibilidad institucional.

**DELGADO FIGUEROA,
NETZAHUALPILLI**

Estudiante de la Maestría de Dirección y Desarrollo de la Innovación por Universidad de Guadalajara, Profesor Universitario en el Centro Universitario de Ciencias de la Salud y Médico Pediatra en Hospital Civil de Guadalajara. Interesado en gestión de la información, síntesis de evidencia científica y tecnología cívica.

DIAS MARTINS, ROBSON

Mestre em biblioteconomia pela UNIRO. Possui especialização em Gestão e Produção Cultural pela Universidade Estácio de Sá. Possui graduação em

	<p>Biblioteconomia pela Universidade Federal do Estado do Rio de Janeiro (2000). Atualmente é bibliotecario do Instituto Nacional de Câncer (INCA). Tendo atuado anteriormente na Universidade do Estado do Rio de Janeiro. Tem experiência na área de Ciência da Informação, com ênfase em Biblioteconomia, atuando principalmente no seguinte tema: Formação e desenvolvimento de coleções</p>
ESCOBAR FARFÁN, JONATHAN ISRAEL	<p>Bibliógrafo en la BDCV, asignado al Centro de Estudios Lingüísticos y Literarios de El Colegio de México. Licenciado en Bibliotecología y Estudios de la Información por la Universidad Nacional Autónoma de México; Master of Arts y Doctor of Philosophy por la School of Languages and Cultures, University of Sheffield. Sus áreas de interés son el patrimonio documental y la cultura libresca, los procesos y herramientas para la edición y el análisis digital, la bibliometría, la diversidad lingüística, la escritura en lenguas minorizadas, el grupo lingüístico náhuatl, y el análisis morfológico.</p>
ESCOBAR VALLARTA, CLAUDIA	<p>Maestra y Licenciada en Bibliotecología y Estudios de la Información (UNAM). Personal académico de la Biblioteca Daniel Cosío Villegas de El Colegio de México desde 2008, donde se ha desempeñado como catalogadora/clasificadora; bibliógrafa para estudios demográficos, urbanos y ambientales; gestora de comunidades virtuales y coordinadora de gestión de colecciones. Ha publicado artículos y capítulos de libro, siendo el más reciente «From Academic Plagiarism to Information Literacy: Mediation in the Ethical Use of Information» (coautoría con Máximo Domínguez) en la <i>Information Literacy in the Workplace: 5th European Conference on Information Literacy (ECIL 2017)</i>, Springer, 2018. Colabora como indizadora del Hispanic American Periodical Index (HAPI) de la UCLA.</p>
FERNANDES DE QUEIROZ, CLAUDETE	<p>Mestre em História, Política e Bens Culturais pela Fundação Getúlio Vargas-RJ. Possui especialização em Docência Superior pelo ISEP (RJ) e graduação em Biblioteconomia pela Universidade Santa Úrsula (RJ). Atuou como Bibliotecária nas seguintes instituições: SENAC/Departamento Nacional; SENAI/RJ/Centro de Tecnologia Euvaldo Lodi; Documentar; Conselho Federal de Enfermagem; Ministério da Defesa/Centro Tecnológico do Exército; e atualmente exerce o cargo de Tecnologista em Saúde Pública na Fiocruz, atuando na coordenação técnica do Repositório Institucional Arca.</p>

GABRIEL JUNIOR, RENE FAUSTINO Graduado em Biblioteconomia e Documentação pela Pontifícia Universidade Católica do Paraná (2008), com mestrado em Ciência, Gestão e Tecnologia da Informação pela Universidade Federal do Paraná (2011) e doutorado em Ciência da Informação pela Universidade Estadual Paulista Júlio de Mesquita Filho (2014). Atualmente é professor adjunto da Universidade Federal do Rio Grande do Sul e do Programa de Pós-Graduação em Ciência da Informação (PPGCIN) da mesma universidade. Tem experiência na área de Ciência da Informação, com ênfase em Biblioteconomia, atuando principalmente nos seguintes temas: Ciência da Informação, Estudos Métricos da Informação, Bibliometria, Brapci, Comunicação Científica, Dados de Pesquisa e Produção Científica. Implantou e coordena a Base de Dados de Periódicos em Ciência da Informação (BRAPCI). Membro do Grupo de Pesquisa de Comunicação Científica e do Núcleo de Estudos em Ciência, Inovação e Tecnologia da UFRGS.

GARCÍA GUERRERO, MONTSERRAT Estudiante del Doctorado en Formación en la Sociedad del Conocimiento de la Universidad de Salamanca (España) y del Doctorado en Gestión Educativa y Política Pública de la Universidad Autónoma de Zacatecas (México). Maestra en Ciencias Sociales y Humanidades con especialidad en comunicación por la Universidad Nacional de Quilmes (Argentina). Licenciada en Letras por la Universidad Autónoma de Zacatecas. Fue coordinadora editorial de la revista *Migración y Desarrollo* por más de diez años y ha participado en más de veinte proyectos de investigación y en el manejo de una plataforma de acceso abierto para las revistas de la Universidad Autónoma de Zacatecas. Especialista en ciencia abierta y apoyo a la producción científica. Gestora de Caxcán Repositorio Institucional de la Universidad Autónoma de Zacatecas y de la Oficina de Ciencia Abierta de la misma Universidad. Miembro fundador de la Red Ciencia Abierta y Repositorios Institucionales (CAyRI) A. C., es parte de la Federación Latinoamericana de Semiótica y de la Asociación Mexicana de Estudios del Discurso. Cuenta con presentaciones en eventos nacionales e internacionales, así como publicación de dos libros como coordinadora, capítulos de libro y de artículos en revistas arbitradas e indexadas.

GARRO ACÓN, MEILYN De profesión Comunicadora audiovisual. Coordinadora del Repositorio Institucional Kérwá de la Universidad de Costa Rica. Líder público de Creative Commons. Miembro de la Subcomisión de Conocimiento Abierto del Consejo Nacional de Rectores y miembro del equipo Latindex Costa Rica. Trabaja desde la Unidad de Gestión de la Calidad de la Vicerrectoría de Investigación en la Universidad de Costa Rica en temas relacionados con cultura libre, ciencia abierta y mejoramiento de la calidad en la distribución de conocimiento.

GÓMEZ RAZZA, VÍCTOR HUMBERTO Ingeniero de Sistemas, CENTRUM MBA Candidate, Project Management Professional (PMP) con amplia experiencia gestionando proyectos de TI. Subdirector de Gestión de la Información y Conocimiento y Asesor en Gestión de Tecnologías del Conocimiento en CONCYTEC.

GONZÁLEZ DÍAZ, GUADALUPE JEANETTE	Coordinadora de la Maestría en Dirección y Desarrollo de la Innovación en la Universidad de Guadalajara.
GUERRERO SOSA, JARED DAVID TADEO	Maestro en Ciencias de la Computación por la Universidad Autónoma de Yucatán (México). Ingeniero en Tecnologías de la Información y Comunicaciones por el Instituto Tecnológico de Chetumal (México). Sus intereses de investigación se enfocan en temas relacionados con los repositorios digitales, la representación del conocimiento y la informática educativa.
GUTIÉRREZ DE LA TORRE, SILVIA EUNICE	Maestra en Humanidades Digitales por la Universidad de Wurzburg y el King's College London. Es Bibliotecaria de Humanidades Digitales de la BDCV, aunque actualmente está en permiso académico para hacer su doctorado en el Grupo de Humanidades Computacionales de la Universidad de Leipzig. Su proyecto actual versa sobre el uso de técnicas de minería de texto y el aprendizaje de máquina para la extracción de referencias de tesis electrónicas.
KLANOVICZ FERREIRA, MANUELA	Licenciada en Ciencias de la Computación (2008), Maestría en Ciencias de la Computación (2012) y estudiante del Doctorado en Comunicación, todos por la Universidade Federal do Rio Grande do Sul. Se incorporó a la Universidad como servidor en 2009, donde desempeña sus actividades como coordinadora del equipo directivo de Lume - Repositorio Digital, en el Centro de Procesamiento de Datos de la Universidad, donde actualmente también participa en el proyecto LUMEdata para la creación de un piloto del Repositorio de Datos de Investigación en la Universidad.
LIRA, ARIEL JORGE	Licenciado en Informática (UNLP). Coordinador informático en SEDICI y CIC-Digital. Integrante del Comité de Expertos del Sistema Nacional de Repositorios Digitales del Ministerio de Ciencia y Tecnología de la Nación (Argentina).
LISOWSKA NAVARRO, MALGORZATA	Directora del CRAI en la Universidad del Rosario (Colombia). Magister en Bibliotecología e Información Científica, Universidad Jagiellona de Cracovia (Polonia). Especialista en Administración de Empresas, Universidad del Rosario. Especialista en Gerencia y Gestión Cultural, Universidad del Rosario. Amplia experiencia en bibliotecas públicas y universitarias, con énfasis en gestión y evaluación bibliotecaria y en implementación de nuevas tecnologías. investigadora en el proyecto de la creación de la Biblioteca Digital Colombiana (BDCOL), coordinación de proyectos internacionales como CoLaBoRa (Comunidad Latinoamericana de Bibliotecas y Repositorios Digitales) y en LA Referencia, patrocinado por de la RedClara y el BID.
LIZARAZO RIVERO, MARÍA LUCÍA	Profesional de recursos de investigación del CRAI en la Universidad del Rosario (Colombia). Magíster en Ingeniería industrial e Ingeniera Industrial de la Universidad Industrial de Santander. Especializada en transferencia de tecnología, gestión de la innovación y con amplio conocimiento en propiedad intelectual, ciencia abierta y gestión de datos de investigación. Con experiencia en la formulación y gestión de proyectos de ciencia, tecnología e innovación.

LÓPEZ ORTIZ DE ARTIÑANO, IANKO	Director del Consorcio Madroño, integrado por siete universidades públicas españolas (Complutense, Autónoma, Politécnica de Madrid, Alcalá, Rey Juan Carlos, Carlos III, UNED). Sus usuarios finales son investigadores, y su objetivo principal es brindar recursos de información a los investigadores y realizar proyectos complejos orientados a mejorar los procesos, herramientas y/o desempeño de las bibliotecas de las universidades asociadas. Tareas principales: definición y gestión estratégica, gestión de proyectos, negociación y gestión de proveedores/vendedores, gestión presupuestaria y financiera, gestión y contratación de recursos humanos, marketing y comunicaciones, relaciones institucionales, preparación y realización de seminarios.
MALDINI, JUAN	Licenciado en Bibliotecología y maestrando en Información y Comunicación por la Universidad de la República (Uruguay). Gerente de Servicios Digitales de la Agencia Nacional de Investigación e Innovación. Coordina SILO, el sistema nacional de repositorios abiertos de ciencia y tecnología de Uruguay.
MANZUR, EZEQUIEL	Analista Programador Universitario y alumno avanzado de la carrera de Licenciatura en Sistemas (UNLP). Es ayudante alumno en las cátedras de Conceptos de Algoritmos, Datos y Programas y Taller de programación de la Facultad de Informática (UNLP). Trabaja en PREBI-SEDICI y CESGI en el proyecto de visibilidad web institucional y en el desarrollo de herramientas de integración de portales web con repositorios digitales.
MARTINS DA COSTA FERREIRA, TIAGO	Atualmente é Sócio-Diretor da Neki IT. Tem experiência na área de Ciência da Computação, com ênfase no desenvolvimento de ferramentas para Saúde Suplementar. Também possui experiência no desenvolvimento de acervos digitais e repositórios institucionais.
MATAS, LAUTARO	Secretario Ejecutivo y Técnico de la Red Latinoamericana de Repositorios de Ciencia Abierta - LA Referencia. Entre 2003 y 2019 formó parte del equipo técnico Red Iberoamericana de Indicadores de Ciencia y Tecnología (RICYT), coordinando el desarrollo de herramientas de recolección, procesamiento y publicación de indicadores de Ciencia, Tecnología e Innovación. Entre 2009 y 2019 fue parte del equipo técnico del Observatorio Iberoamericano de la Ciencia, la Tecnología y la Sociedad (CAEU/OEI), actuando como Coordinador/Desarrollador del portal Intelligo. Entre los años 2005 y 2011, se desempeñó en el Centro Argentino de Información Científica y Tecnológica (CAICYT-CONICET) al frente del proyecto SciELO Argentina y en desarrollo de software para área de patentes e inteligencia estratégica. En 2013 fue seleccionado como consultor/desarrollador del piloto de software de cosecha del proyecto BID que dio origen a la plataforma tecnológica LA Referencia, ejecutado por la Cooperación Latino Americana de Redes Avanzadas (CLARA) y los organismos de Ciencia y Tecnología nacionales asociados al proyecto.

MENÉNDEZ DOMÍNGUEZ, VÍCTOR HUGO	Doctor en Tecnologías Informáticas Avanzadas por la Universidad de Castilla-La Mancha (España). Es Profesor Titular en la Facultad de Matemáticas de la Universidad Autónoma de Yucatán (México). Su trabajo de investigación se centra en temas relacionados con repositorios de documentos digitales, la representación del conocimiento y la gestión de objetos de aprendizaje.
MORA CAMPOS, ANDREA MARÍA	Asesora académica para Ciencia Abierta (revistas científicas y repositorios, gestión de datos de investigación) de la Rectoría Adjunta y de la Vicerrectoría de Investigación de la Universidad Nacional (2015 a la actualidad). Del 2015 a la actualidad miembro de la Subcomisión de Conocimiento Abierto de CONARE, representante de la Universidad Nacional (Costa Rica). También desde el 2017 a la actualidad es la representante de Costa Rica para LA Referencia. Miembro fundador del nodo RDA-Costa Rica. Miembro del grupo centroamericano DSpace. Miembro del grupo VIVO en español. Fue editora en jefe de varias revistas científicas de la Universidad Nacional y Universidad de Costa Rica. Cuenta con varias publicaciones y ha sido expositora y promotora internacional en temas de Ciencia Abierta.
MORÁS SILVA, LEILA	Graduada em Biblioteconomia pela Universidade Federal do Rio Grande do Sul (2018) e Pós-Graduada em Gestão de Documentos e Informação (2020). Atualmente é Mestranda do Programa de Pós-Graduação em Ciência da Informação (PPGCIN-UFRGS). Tem experiência na área de Ciência da Informação, Biblioteca Universitária e Biblioteca Pública. Área de interesse: Comunicação Científica, Produção Científica, Ciência da Informação.
OLIVEIRA SIMÕES, KÁTIA DE	Mestre em Biblioteconomia pela Universidade Federal do Estado do Rio de Janeiro (UNIRIO), Especialista em Informação Científica e tecnológica em saúde, pela Fundação Oswaldo Cruz (FIOCRUZ). Com experiência na área de Ciência da Informação, indexação temática para recuperação de Informação, Repositório Institucional, treinamento de usuário na área de saúde. Principais temas de atuação: Acesso à informação, Bibliotecas virtuais, Gestão da informação, Serviços de Informação.
ORTIZ REYES, JOSÉ VALENTÍN	Licenciado en Comunicación Social por la UAM Xochimilco, maestro en Bibliotecología por El Colegio de México y doctor Bibliotecología y Estudios de la Información por la UNAM. Se ha desempeñado como bibliógrafo y como Coordinador de servicios en la BDCV. Sus temas de interés son comportamiento informativo, formación de usuarios, teoría fundamentada y periodismo.
PORT DA ROCHA, RAFAEL	Profesor Asociado del Departamento de Ciencias de la Información de la Universidad Federal de Rio Grande do Sul (UFRGS) (Brasil), actuando en programas de pregrado en Archivística y Biblioteconomía y de posgrado en Ciencias de la Información. Fundador y subdirector del Centro de Documentación y Colecciones Digitales de Investigación (CEDAP) de la UFRGS. Coordinador del GT-RDP/UFRGS, grupo

	<p>de trabajo conducido por RNP (Rede Nacional de Ensino e Pesquisa, Brasil) para identificar prácticas, analizar requerimientos y desarrollar prototipos de repositorios de datos de investigación. Graduación, maestría y doctorado en informática por la Universidad Federal de Rio Grande do Sul, Brasil. Desarrolla investigación en ontologías, metadatos, repositorios digitales, preservación digital, datos de investigación y digitalización de colecciones.</p>
QUIROA HERRERA, MARÍA LOURDES	<p>Doctora en Documentación por la Universidad Carlos III de Madrid (España). Bibliógrafa para el Centro de Estudios Internacionales. Ha impartido el Curso de Investigación Documental en Ciencias Sociales y Humanidades en diferentes Centros de El Colegio. Sus áreas de interés son servicios de referencia, referencia virtual, alfabetización informacional.</p>
RAZO RODRÍGUEZ, ANTONIO FELIPE	<p>Ingeniero en Sistemas Computacionales, Maestro en Ciencias con Especialidad en Ingeniería en Sistemas Computacionales y Maestro en Diseño de Información por la Universidad de las Américas Puebla (UDLAP). Desde el 2022 es Director de Bibliotecas de la UDLAP. Responsable del desarrollo y mantenimiento técnico de la Red Mexicana de Repositorios Institucionales (REMERI) desde el 2012. Consultor para la primera etapa del proyecto LA Referencia del Banco Interamericano de Desarrollo y RedCLARA en el 2013. Desde el 2016 a la fecha, colaborador independiente en proyectos de Repositorios Institucionales con DSpace con las siguientes instituciones en México: UASLP, UTM, UAZ, UAT, UAN, UPPue, UP, ITESM, LANIA, UDLAP, CONACyT y CUDI. Desde el 2020 consultor en DSpace y DSpace CRIS para Scimago Research Group (Colombia) para instituciones en México, Costa Rica, República Dominicana y Colombia.</p>
RIVEIRO DE CARVALHO SEGUNDO, WASHINGTON LUIS	<p>Tiene un doctorado y una maestría en informática de la Universidad de Brasilia, con una pasantía de doctorado Sandwich en el King's College de Londres. Graduado en Matemáticas, también de la Universidad de Brasilia. Es Coordinador Técnico del Área de Tratamiento, Análisis y Difusión de Información Científica en el Instituto Brasileño de Información en Ciencia y Tecnología (IBICT/MCTI). Es miembro de los comités científicos y organizadores de conferencias: Conferencia de Ciencia Abierta Luso-Brasileña; Conferencia Internacional de Repositorios Abiertos; y Foro internacional sobre la búsqueda de sistemas expertos. Es miembro de los siguientes grupos: Grupo Latinoamericano de Expertos en Datacite; Grupo Asesor de OpenDOAR; Grupo Asesor de ROR; Políticas de datos de investigación de RDA en América Latina para repositorios; Grupo de Coordinación de GoFAIR Brasil; y Grupo de Especialistas en Revistas Científicas de la CPLP. Coordina los Proyectos: Estudios y desarrollo de interoperabilidad entre sistemas de información con miras a la integración y gestión de contenidos técnico-científicos; Funarte-CEDOC; IPHAN-INRC; y Repositorio Funarte-Arte Digital. Es líder del Grupo de Investigación y Laboratorio del Ecosistema de Investigación Científica Brasileña (LaEPeCBr)</p>

RODRIGUES, ELOY	Director de la Universidade do Minho - Unidade de Serviços de Documentação e Bibliotecas (USDB). Dirige y es líder de UMINHO además de liderar el equipo técnico que está desarrollando el RCAAP (Repositorio Científico de Acceso Abierto de Portugal) y ha participado en varios proyectos relacionados con repositorios de acceso abierto.
ROMERO MILLÁN, CAMELIA	Doctora en Bibliotecología y Estudios de la Información por la UNAM. Especialista en fuentes de información y lenguaje controlado en Estudios de Género. Bibliógrafa para el Centro de Estudios Sociológicos y el Centro de Estudios de Género. Profesora del curso de Investigación Documental a grupos de Maestría en Sociología y Estudios de Género de El Colegio de México. Organizadora del sitio web Género en México, fuentes para su historia. Ha participado en foros como la Asociación de Estudios Latinoamericanos (LASA), Seminar on the Acquisition of Latin American Library Materials (SALALM) y la International Society of Knowledge Organization (ISKO).
ROZENBLUM, CECILIA	Coordinadora General de Revistas Científicas de la UNLP, Facultad de Humanidades y Ciencias de la Educación, Universidad Nacional de La Plata. Licenciada en Bibliotecología y Documentación (UNLP), Magister en Ciencia, Tecnología y Sociedad (UNQ). Desde 2005 se especializa en el asesoramiento y coordinación de la gestión, edición e indización de revistas científicas, especialmente en las áreas de ciencias sociales y humanidades. Desde 2012 a 2018 fue Directora de Gestión Editorial de Publicaciones Periódicas de la Facultad de Humanidades y Ciencias de la Educación (UNLP). Entre 2018 y 2021, coordinadora general de revistas de la UNLP y desde junio asesora general. Desde 2014 es Adjunta ordinaria a cargo de la Cátedra Servicio de información y Referencia Especializada de la carrera de Bibliotecología de la misma institución.
SANTOS, GILDENIR CAROLINO	Bacharel em Biblioteconomia pela Pontifícia Universidade Católica de Campinas. Mestre e Doutor em Educação pela Universidade Estadual de Campinas (UNICAMP). Pós-doutor em Divulgação Científica pela UNICAMP. Gestor do Portal de Periódicos Eletrônicos Científicos da Universidade Estadual de Campinas (UNICAMP). Pesquisador do Grupo de Pesquisa Estudos e Práticas de Preservação Digital – Dríade (Rede Cariniana/IBICT).
SIFUENTES ARROYO, RAÚL	Más de veinte años de experiencia en automatización integral de centros de información y bibliotecas universitarias, implementación de repositorios institucionales, bases de datos y revistas electrónicas. Consultor internacional en tecnología de la información aplicada a bibliotecas en distintas universidades de América Latina por más de quince años. Desde hace cinco años me desempeño como Gestor de Tecnología e Innovación en el Sistema de Bibliotecas de la PUCP. Licenciado en Bibliotecología y Ciencia de la Información por la Pontificia Universidad Católica del Perú; Magíster en Gestión y Política de la Innovación y la Tecnología por la Pontificia Universidad Católica del Perú.

SILVA, CAMILA MARIANA APARECIDA DA	Possui graduação em Biblioteconomia e História pela Universidade Federal de Minas Gerais (UFMG). Doutoranda e mestre pelo Programa de Pós-Graduação em Ciência da Informação pela Escola de Ciência da Informação (UFMG). Bibliotecária integrante da equipe gestora do Repositório Institucional da UFMG.
SIMIÃO CARDOZO, JULIANO	Estudante de graduação de Medicina pela Faculdade de Ciências Médicas da Universidade Estadual de Campinas (UNICAMP). Bolsista-trabalho no Portal de Periódicos da Universidade Estadual de Campinas (UNICAMP).
SUCAR SUCCAR, ENRIQUE	Estudió Ingeniería Electrónica y Comunicaciones en el ITESM Campus Monterrey; obtuvo la maestría en Ingeniería Eléctrica por la Universidad de Stanford, California, y el doctorado en Computación por el Imperial College de Londres. Ha realizado estancias de investigación en el Imperial College, en la Universidad de British Columbia (Canadá), en el INRIA (Francia), y en CREATE-NET (Italia). Tiene experiencia como investigador y profesor en el Instituto de Investigaciones Eléctricas, el Tecnológico de Monterrey y actualmente trabaja en la Coordinación de Ciencias Computacionales del Instituto Nacional de Astrofísica, Óptica y Electrónica de Puebla. Es miembro del Sistema Nacional de Investigadores, de la Academia Mexicana de Ciencias, de la Academia de Ingeniería, miembro Fundador de la Academia Mexicana de Computación y Senior Member IEEE. Recibió el Premio TECNOS 2014 por el desarrollo del sistema Gesture Therapy para la rehabilitación de pacientes con embolia cerebral y en 2016 el Premio Nacional de Ciencias en el área de Tecnología, Innovación y Diseño. Ha sido presidente de la Sociedad Mexicana de Inteligencia Artificial, de la Federación Mexicana de Robótica y de la Academia Mexicana de Computación. Editor asociado de las revistas <i>Computación y Sistemas</i> , <i>Pattern Recognition</i> y <i>Computational Intelligence</i> . Su investigación se centra en entender y modelar las capacidades mentales que nos permiten razonar y tomar decisiones bajo incertidumbre y desarrollar sistemas inteligentes basados en programas de software.
SUMBA, FREDDY	Ingeniero en Sistemas Computacionales, especialista en gestión de productos de software y Google Cloud computing certified. Trabaja en el área de I+D+i en CEDIA. Dedicado a la creación de nuevos servicios tecnológicos para la comunidad académica del Ecuador y la región. Actualmente está trabajando en proyectos como la Red de Repositorios de Acceso Abierto del Ecuador (RRAAE), el Catálogo de servicio tecnológicos (RETEC), la plataforma de investigación basada en CRIS-REDI+ y actualmente en el proyecto The Dublin Core Scholarly Resources Application Profile SRAP con Dublín Core Metadata Initiative. Activista de ciencia abierta y apasionado en tecnopolítica, ha colaborado en varios proyectos en función del desarrollo de proyectos basados. Colabora desde CEDIA con diferentes instituciones académicas y científicas en el desarrollo de políticas, tecnologías y mejores prácticas para la adopción de repositorios académicos en Ecuador.

TENA ESPINOZA DE LOS MONTEROS, MARTÍN ADALBERTO	Doctorando del Doctorado en Formación en la Sociedad del Conocimiento de la Universidad de Salamanca (USAL), Master en Bibliotecología y Ciencias de la Información por el Instituto Tecnológico y de Estudios Superiores de Monterrey (ITESM) (2006). Bibliotecario gestor de proyectos en el Sistema Universitario de Bibliotecas de la Universidad de la Universidad de Guadalajara (SiUBiUDG). Profesor en el Sistema de Universidad Virtual (UDGVirtual) de la Universidad de Guadalajara. Colaborador y promotor de proyectos e iniciativas a favor del acceso abierto, la ciencia abierta y el conocimiento libre.
TETTAMANTI, SANTIAGO	Analista Programador Universitario, egresado de la Facultad de Informática de la UNLP. Actualmente es alumno avanzado en la carrera de Licenciatura en Informática y se desempeña como ayudante alumno en las cátedras de Programación Concurrente y Sistemas Distribuidos y Paralelos de la misma facultad. Desde fines de 2017 integra el equipo de desarrollo de los repositorios SEDICI y CIC-Digital en PREBI-SEDICI (UNLP).
TEXIER, JOSÉ DANIEL	Doctor e Ingeniero, docente-investigador de la Universidad Nacional de Chilecito (Argentina). Autor de varios artículos científicos en revistas indexadas en las bases de datos Scopus, WoS, SciELO y Redalyc. Ha participado en eventos nacionales e internacionales como ponente, moderador y comités científicos en Estados Unidos, Colombia, Ecuador, Costa Rica, México, Venezuela y Argentina. Su formación académica está centrada en Ciencias de la Computación. Graduado como Ingeniero Informático de la Universidad Nacional Experimental del Táchira (Venezuela), luego como Magíster en Computación de la Universidad de los Andes (Venezuela). Después como Doctor en Ciencias Informáticas en la Universidad Nacional de La Plata (Argentina), con tesis doctoral sobre las Ciencias de la Computación aplicadas en las Ciencias de la Información y Documentación. Sus líneas de investigación están enmarcadas en bibliotecas digitales/repositorios institucionales, compiladores e intérpretes y procesamiento de lenguaje natural.
URIBE TIRADO, ALEJANDRO	Doctorado por la Universidad de Granada en Bibliotecología y Documentación Científica, con una maestría en Ingeniería por la Universidad Eafit (Colombia) y especializaciones en Gerencia de Servicios de Información por la Universidad De Antioquia (Colombia) y Negocios Electrónicos por la Escuela Española de Negocios Internacionales. Actualmente se desempeña como Profesor-Investigador en la Escuela Interamericana de Bibliotecología,. Universidad de Antioquia (Colombia).
VÁZQUEZ TAPIA, ROSALINA	Ingeniero en Sistemas Computacionales y Maestra en Tecnología Educativa, actualmente cursa el Doctorado en Formación de la Sociedad del Conocimiento de la Universidad de Salamanca (España). Desde 1992 es funcionaria e investigadora de la Universidad Autónoma de San Luis Potosí (México), ocupando diferentes puestos directivos. Es Fundadora y Coordinadora General de la Red Mexicana de

Repositorios Institucionales (REMERI) y Responsable Técnica del nodo técnico de México en LA Referencia desde el año 2012. Ha sido evaluadora y responsable técnica de proyectos en convocatorias internacionales y de CONACYT, y autora de múltiples trabajos, ponencias y conferencias en eventos académicos. Además, colabora con las siguientes organizaciones e iniciativas nacionales e internacionales en materia de acceso abierto, ciencia abierta, tecnologías educativas y patrimonio documental: miembro del Comité permanente de la Conferencia Internacional BIREDIAL-ISTEC; socio técnico del Proyecto Los Primeros Libros de las Américas (PLA); miembro de la Confederación de Repositorios (COAR) y Organización Internacional EuroCRIS; coordinadora de la Comisión de Repositorios y Recursos Educativos Digitales del Grupo de Tecnología Educativa del Comité ANUIES-TIC; coordinadora del Eje de Tecnología Educativa y Servicios digitales del Equipo Colaborativo para la Transformación Digital de la Educación (EduTraDi).

VILA, MARÍA MARTA

Analista Programadora Universitaria, integra desde el año 2003 el staff permanente del repositorio central digital de la Universidad Nacional La Plata (UNLP). Dentro del equipo de trabajo de SEDICI realiza tareas de desarrollo, investigación y mantenimiento de software. Los principales lineamientos de su labor se relacionan con el web scraping, harvesting, OAI-PMH, visibilidad Web, interoperabilidad de repositorios y sistemas de gestión de contenidos, entre otros. Desde el año 2016 participa además del grupo de preservación digital, difusión y análisis de impacto del Centro de Servicios en Gestión de Información (CESGI) de la Comisión de Investigaciones Científicas (CIC) de la provincia de Buenos Aires.

VILLARREAL, GONZALO LUJÁN

Actividad de I+D+T: Miembro de PREBI-SEDICI desde el año 2004, coordinador del Portal de Revistas (2008), Portal de Congresos (2009), Proyecto de Visibilidad Web Institucional (2012) y Portal de Libros (2015). Coordinador del desarrollo del software Celsius, y responsable de la gestión de la infraestructura informática del Consorcio Iberoamericano para la Producción en Ciencia y Tecnología (ISTEC). Subdirector del Centro de Servicios en Gestión de Información (CESGI, 2016) de la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires Principales áreas de trabajo: métricas de visibilidad e impacto de la producción científica y académica; interoperabilidad y visibilidad de la producción científica; sistemas de producción y publicación: portales de revistas, servicios editoriales, soporte a congresos, formatos de publicación; infraestructura y gestión de servidores. Actividad docente: Jefe de Trabajos Prácticos de la materia Conceptos de Algoritmos, Datos y Programas y de la materia Taller de Programación (Facultad de Informática, UNLP). Profesor del curso de ingreso en la Facultad de Informática en sus distintas ediciones. Docente en los cursos de posgrado Bibliotecas y Repositorios Digitales. Tecnología y Aplicaciones y Medida, visibilidad e Impacto de la producción científica, ambos de la carrera de Doctorado en Ciencias Informáticas (UNLP). Docente en cursos de posgrado en otras unidades académicas. Tallerista y conferencista en eventos científicos,

académicos y de divulgación.

VILLEGAS TOVAR, RICARDO

Máster en Gestión de la Información y el conocimiento egresado de la Universitat Oberta de Catalunya (España). Cuenta con el doble grado de maestría en Derecho de la Propiedad Intelectual otorgado por la Organización Mundial de la Propiedad Intelectual y la Universidad de Turín (Italia). En 2019 defendió su tesis doctoral en Ciencias de Gobierno en la que hizo un estudio de los aspectos de calidad que deben ser considerados en la gestión de los datos abiertos. Tiene una trayectoria de veinte años en el medio bibliotecario universitario, principalmente en el área de servicios especializados de información para las áreas de investigación científica. En 2017 se sumó a la Vicerrectoría de Investigación y Estudios de Posgrado en la Benemérita Universidad Autónoma de Puebla. Ahí se desempeña como Coordinador de Producción Académica y Visibilidad Internacional.

VINDAS SOLANO, SOFÍA

Posee un Doctorado en Historia por la Universidad de Costa Rica, en su tesis doctoral investigó sobre la consolidación de los museos de arte moderno en Guatemala y Costa Rica. Obtuvo un Máster en Ciencias Políticas, así como un Bachillerato en Historia e Historia del Arte en la UCR. Actualmente es docente de la Escuela de Historia, y es investigadora en el Instituto de Investigaciones en Arte, donde coordina el Repositorio Centroamericano de Patrimonio Cultural.



ESTE LIBRO DE ACTAS TERMINÓ DE COMPONERSE EN ABRIL DE 2022.

PARA SU COMPOSICIÓN SE UTILIZÓ LA FAMILIA DE FUENTES ROBOTO

(ROBOTO, ROBOTO BLACK Y ROBOTO THIN).

