

Organizan



ACTAS

CONFERENCIA INTERNACIONAL

# BIREDDIAL-ISTEE

18-20 de octubre de 2023

MONTEVIDEO • URUGUAY



Co-organizan



Red Mexicana  
de Repositorios  
Institucionales



---

Patrocinadores oro

**ACCUCOMS**

**ANII** AGENCIA NACIONAL DE INVESTIGACION E INNOVACION

**SPRINGER NATURE**  Wolters Kluwer

---

Patrocinadores plata

 **Clarivate™**

PROTECTORADO DE GESTION   UNIVERSIDAD DE LA REPUBLICA URUGUAY

---

Apoyan





XII CONFERENCIA INTERNACIONAL DE  
BIBLIOTECAS Y REPOSITARIOS DIGITALES

18 al 20 de octubre de 2023

SEDE

MONTEVIDEO (URUGUAY)

Actas de la Conferencia Internacional BIREDIAL-ISTEC 2023 / Fernanda Beigel ...  
[et al.] ; Compilación de Marisa De Giusti ; Coordinación general de Marisa De  
Giusti ; Editado por Analía V. Pinto. - 1a ed compendiada. - La Plata :

Universidad

Nacional de La Plata. SEDICI, 2024.

Libro digital, PDF

Archivo Digital: descarga y online

ISBN 978-950-34-2375-2

1. Bibliotecología. 2. Recursos Informáticos. I. Beigel, Fernanda II. De Giusti,  
Marisa, comp. III. Pinto, Analía V., ed.

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

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





# Índice

Comité Organizador Local _____	8
Comité Científico _____	8
Sede 2023 _____	9
Mesas y talleres _____	10
Conferencias magistrales _____	12
<b>El proyecto de ciencia abierta en un mundo desigual: avances, desafíos y tensiones</b> _____	13
<b>WorldFAIR: Open Science and FAIR Data for Cross-Domain Grand Challenges</b> _____	15
<b>CoARA: una coalición, otra iniciativa o la oportunidad definitiva para reformar la evaluación de la investigación</b> _____	17
<b>Infraestructuras, plataformas y tecnologías del Ecosistema Regional de Ciencia Abierta</b> _____	18
Ponencias _____	19
<b>Alianza SIDALC: plataforma hemisférica que promueve el acceso abierto a información agropecuaria</b> _____	20
<b>Automatización de la evaluación de repositorios institucionales</b> _____	32
<b>O desafio do acesso aberto para os povos originários Kaingang na era digital</b> _____	55
<b>Rede Sudeste de Repositórios Digitais: a importância do trabalho colaborativo</b> _____	73

Accesibilidad de artículos en formato PDF en revistas científicas de acceso abierto _____	88
Disponibilização de dados de pesquisa em repositórios universitários da América Latina _____	101
El rol de la biblioteca universitaria en la gestión de datos de investigación: una estrategia para potenciar la visibilidad científica y la ciencia abierta _____	127
Criptografia de arquivo em Biblioteca Digital com o software Dspace _	145
Analysis of the Open Data Landscape in Mexico _____	165
Análisis de escenarios y protocolos para integración de repositorios digitales y sitios web institucionales _____	181
Interface do gestor: uma dashboard para gestores de repositórios digitais _____	197
Vitrinas del conocimiento: una oportunidad para la acción _____	215
Avaliação de acessibilidade digital: Biblioteca Digital Brasileira de Teses e Dissertações (BDTD) _____	223
Criação de Metadados Dublin Core para documentos de patentes no Arca – Repositório Institucional da Fundação Oswaldo Cruz _____	241
Dados abertos para a sustentabilidade: avaliação usando os princípios FAIR _____	254
Competência em Dados de Pesquisa: analisando as práticas dos profissionais da informação envolvidos no apoio à pesquisa _____	256
Criação de um chatbot para o ARCA - Repositório Institucional da Fiocruz _____	276
A aplicação da metodologia das boas práticas na submissão e autoarquivamento de recursos educacionais abertos em repositórios_	286

Autoavaliação de confiabilidade do Arca – Repositório Institucional da Fiocruz pela norma ISO-16363:2012: apresentação dos resultados consolidados _____	301
Formación de competencias para la Ciencia Abierta en la Universidad del Rosario (Colombia) _____	318
Acesso Aberto Verde: breves reflexões a partir de casos da produção intelectual publicada por Instituições de Ensino brasileiras _____	333
e-compartilhando _____	352
Práticas de compartilhamento de dados pelos pesquisadores PQ da área de Física e Astronomia do CNPQ _____	354
Geração de Valor a partir da Oferta de Repositórios Institucionais de Dados de Pesquisa _____	356
Modelo dimensional para la medición de la producción académica ____	371
Interoperabilidad automática gracias al uso de PIDs - ORCID _____	390
Taxonomía de la ciencia abierta: revisada y ampliada _____	391
Pósters _____	394
Rede Sudeste de Repositórios Digitais: a importância do trabalho colaborativo _____	395
Analysis of the Open Data Landscape in Mexico _____	397
Análise da viabilidade de associação dos planos de gestão de dados acionáveis por máquina com o ciclo de vida dos dados no ecossistema de pesquisa brasileira _____	400
Infraestructura para soportar la Ciencia Abierta en la Universidad del Rosario _____	403
Periódicos científicos: normativas para abertura de dados _____	405

El impacto de los Identificadores Persistentes (PIDs) en el Repositorio de Datos Académicos RDA-UNR _____	407
Práticas de compartilhamento de dados pelos pesquisadores PQ da área de Física e Astronomia do CNPQ _____	410
Interoperabilidad automática gracias al uso de PIDs - ORCID _____	412
Taxonomía de la ciencia abierta: revisada y ampliada _____	414

## Comité Organizador Local

CARINA PATRÓN, Mesa de Directoras de Bibliotecas, Universidad de la República

CHIARA MIRANDA, Facultad de Información y Comunicación, Universidad de la República

GREGORY RANDALL, Facultad de Ingeniería, Universidad de la República

JUAN MALDINI, Agencia Nacional de Investigación e Innovación

MABEL SEROUBIÁN, Servicio Central de Informática Universitaria, Universidad de la República

MAGELA CABRERA, Facultad de Información y Comunicación, Universidad de la República

VALENTINA MORANDI, Universidad de Montevideo

EMILENA CARDOZO, Asociación Uruguaya de Archivólogos

ESTHER PAILOS Y LAURA SCHELOTTO, Asociación de Bibliotecólogos del Uruguay

SYLVIA PIOVESAN, Asociación Uruguaya de Revistas Académicas

## Comité Científico

CATERINA GROPOSO, Universidade Federal do Rio Grande do Sul (Brasil)

MALGORZATA LISOWSKA, Universidad del Rosario (Colombia)

MARISA R. DE GIUSTI, Universidad Nacional de La Plata (Argentina)

MARLEM URIBE MARENCO, Universidad del Norte (Colombia)

MEILYN GARRO, Universidad de Costa Rica (Costa Rica)

ROSALINA VÁZQUEZ, Red Mexicana de Repositorios Institucionales (México)

## Sede 2023

La **Universidad de la República (UDELAR)** es la universidad pública más grande de Uruguay, con más de 145.000 estudiantes de grado. Fue fundada el 18 de julio de 1849 en Montevideo, y actualmente tiene presencia en 14 departamentos del país. Cuenta con dieciséis facultades y varias escuelas, campus regionales e institutos universitarios.

La **Facultad de Información y Comunicación (FIC)** es una de las 16 facultades de la Universidad de la República, creada en 2013. Se encuentra ubicada en la ciudad de Montevideo, calle San Salvador 1944, barrio Parque Rodó. Ofrece tres carreras de grado: Licenciatura en Archivología, Licenciatura en Bibliotecología y Licenciatura en Comunicación, además de cursos de educación permanente y carreras de posgrado.

FIC es el servicio académico de UDELAR que se propone atender la información y la comunicación como objeto de estudio específico. En cuanto a la información, es el único ámbito educativo del país dedicado a la formación de profesionales en las áreas de bibliotecología y archivología. Respecto a la comunicación, brinda una formación profesional integral orientada a la construcción y análisis crítico-propositivo de los procesos comunicacionales.

# Mesas y talleres

## **MESA: POLÍTICAS NACIONALES DE CIENCIA ABIERTA**

A cargo de: MARISA DE GIUSTI (UNLP, Argentina), RENÉ FAUSTINO GABRIEL JUNIOR (UFRGS, Brasil), ROSALINA VÁZQUEZ (REMERI, México), MARLEM URIBE (UNINORTE, Colombia). Modera: CATERINA GROPOSO PAVÃO (UFRGS).

## **TALLERES**

### **GESTIÓN DE DATOS DE INVESTIGACIÓN**

Dictado por **MALGORZATA LISOWSKA**

Objetivo: Sensibilizar y adquirir los conocimientos sobre la gestión de datos de investigación en una institución de educación superior, teniendo en cuenta las diversas variables y diferentes actores que están involucrados en el proceso. Es un taller presencial, teórico-práctico que emplea el juego de roles e invita a la discusión por parte de todos los participantes, quienes deben lograr su misión secreta. El taller se acompaña con una presentación teórica de los diferentes componentes del plan de gestión de datos, además de resolver dudas que surjan durante su desarrollo por parte de los participantes. El taller tuvo como objetivo lograr la construcción del plan de gestión de datos del proyecto asignado (de diferente índole), y cada participante debió personificar su rol para aportar en la discusión y resolver el reto. Al finalizar las discusiones, se realizó una plenaria en la cual cada grupo compartió los resultados de sus discusiones y los integrantes expusieron sí lograron cumplir o no su misión secreta.

## **PRESERVACIÓN A LARGO PLAZO: RECOMENDACIONES PARA UN REPOSITORIO INSTITUCIONAL**

Dictado por **MARISA DE GIUSTI**

Este workshop estuvo dedicado a relatar las actividades recomendadas en un repositorio institucional para cumplir con su objetivo central de preservar el patrimonio y la producción académica, científica y de extensión de una institución de educación superior. Las actividades de preservación en un repositorio institucional están guiadas por una política, que es el documento de más alto nivel que define los objetivos a cumplir; por debajo de este documento se encuentra el plan de preservación y tras él, y a más bajo nivel, todas las actividades que se realizan en atención a cumplir con la norma ISO 14721, que define las funciones necesarias en un repositorio para cumplir con la obligación de preservar y dar acceso a largo plazo a sus contenidos: los procesos de ingesta, archivo, administración, gestión de datos, planeamiento de la preservación y acceso, así como el seguimiento del paquete de información a lo largo del ciclo de vida. Dentro del devenir de actividades también se mostraron procesos vinculados a una apropiada digitalización dentro de un repositorio institucional para asegurar la producción de documentos preservables y de calidad. Asimismo se hizo un recorrido por la documentación que debería tenerse, en acuerdo con la norma ISO 16363 (de certificación) que es un punto extremadamente débil en todos los RI.

## **GESTIÓN DE DATAVERSE Y BUENAS PRÁCTICAS PARA LA ORGANIZACIÓN DE SETS DE DATOS**

Dictado por **LAURA REZENDE**

El workshop trató de datos científicos abiertos: definiciones y perspectivas de la apertura; conceptos y normativas que implican la gestión de datos científicos a nivel institucional; repositorios de datos científicos que utilizan el software Dataverse: principales funcionalidades, funciones y posibilidades para la conservación de datos; creación y configuración de colecciones; asignación de permisos de usuario; funcionalidades de Dataverse a nivel de comunidades Dataverse, conjunto de datos y archivo; sumisión de datos en la práctica (sumisión de conjuntos de datos y archivos).



BIREDIAL  
ISTEC



Conferencias  
magistrales

# El proyecto de ciencia abierta en un mundo desigual: avances, desafíos y tensiones

**FERNANDA BEIGEL**

(ARGENTINA)

## RESUMEN

La Recomendación de Ciencia Abierta de la UNESCO, aprobada en noviembre de 2021, propone promover un consenso global sobre sus valores y acciones. Los pilares de la apertura que se propone este proyecto son las infraestructuras científicas abiertas, el diálogo con distintos sistemas de conocimiento, así como el compromiso con los distintos sectores sociales, es decir la ciencia ciudadana y participativa. Las cinco manifestaciones principales de esta apertura de la ciencia son el acceso abierto a las publicaciones científicas, el acceso abierto a los datos de investigación, los recursos educativos abiertos, el software y el hardware abiertos. Ahora bien, para crear un entorno normativo propicio para la ciencia abierta, la Recomendación plantea la necesidad de fomentar prácticas responsables en materia de evaluación de la investigación que incentiven la ciencia de calidad, reconociendo la diversidad de los resultados y estimulando las distintas misiones de la universidad. Se interpela a los Estados miembros para que, de acuerdo con sus circunstancias, estructuras de gobierno y disposiciones constitucionales específicas, y conforme a los marcos jurídicos internacionales y nacionales, procuren activamente eliminar los obstáculos para la ciencia abierta, especialmente los relativos a los sistemas de evaluación y el avance de las buenas prácticas en materia de ciencia abierta. Promocionar una cultura de ciencia abierta es una tarea compleja que requiere sistemas integrados de información que permitan conocer, promover y evaluar el universo de producciones y actividades de investigación. Por eso, una de las preocupaciones principales planteadas en esta Recomendación es que, aun con sus buenas intenciones, la ciencia abierta podría amplificar la brecha entre países tecnológicamente más avanzados y los países más pobres, con infraestructura digital precaria. Así, el crecimiento unilateral de plataformas de ciencia abierta en los países dominantes no sólo incrementaría las desigualdades de acceso a la ciencia, sino que habilitaría diferentes formas de exacción y comercialización de datos provenientes de la periferia. Aquí se abordan las desigualdades de infraestructura que afectan a los países de bajos y medianos ingresos, señalando las principales asimetrías que condicionan los caminos de la ciencia abierta en el Sur global. En esta conferencia repasaré la situación de las infraestructuras de ciencia abierta, la dotación mundial de repositorios,

sistemas integrados de información científica y revistas científicas. Luego, discutiré el rol crítico que tienen los sistemas de evaluación de la región para producir una transformación de la magnitud que anida en la ciencia abierta, sin resignar soberanía y anclaje social.

# WorldFAIR: Open Science and FAIR Data for Cross-Domain Grand Challenges

**SIMON HODSON**

(FRANCIA)

## **ABSTRACT**

First published in 2016, in an article that has now received over 6000 citations, the FAIR (Findability, Accessibility, Interoperability, Reusability) Principles have had a significant influence in policy, practice and thinking about research data management and stewardship. The catchy mnemonic has been effective, but the article and related work convey an important message. The fundamental purpose of the FAIR principles is to provide guidelines such that data and metadata relevant to all kinds of research outputs are machine readable and machine actionable. The vision is one in which research outputs can be visited online and at vast scale and the potential of machine assisted analysis can be realised, but with data and metadata that are sufficiently reliable so as to reduce error and quantify uncertainty. Another catchphrase is that FAIR also means Fully AI Ready. The FAIR principles do not insist that data or metadata be fully open, and make the case that the principles apply just as well to cases where data can only be accessed in controlled circumstances. Yet, the interest in FAIR is related to another major development and movement in early 21st century research: that of Open Science. Recognising, of course, that in some circumstances access restrictions are necessary, the Open Science movement aims to maximise the public benefit of research (particularly publicly funded research) by insisting that scientific outputs should be as open as possible, and that access restrictions should be «proportionate and justified». The nuances of these arguments are frequently lost, and sometimes obfuscated by vested interests. Nevertheless, the UNESCO Recommendation on Open Science, supported by the International Science Council, represents another major step in the transformation of science policy. Why does this matter? Why should we care about these developments in high level policy and in data practice? The major global scientific and human challenges of the 21st century (including climate mitigation and adaptation, sustainable development, disaster risk research and reduction, smart cities and energy systems, precision medicine or agriculture) can only be addressed through cross-domain research that seeks to understand complex systems through machine-assisted analysis at scale. Our capacity for such analysis is currently constrained by the limitations in our ability to access and combine heterogenous data within and across domains. The FAIR principles and the frameworks set by Open

Science provide a significant part of the solution. Attention needs to be paid to the interfaces where data is used between disciplines, but the engagement with key research disciplines around data (re-)usability is essential. To help address these issues, CODATA has been entrusted by the International Science Council (ISC) to develop a programme of activity: Making Data Work for Cross-Domain Grand Challenges. After some exploratory work, the flagship activity is the WorldFAIR project which focuses on the implementation of the FAIR principles both within and across 11 different domain and cross-domain case studies, with a central effort to understand and guide cross-domain FAIR. It is the first broad-based effort to understand the issues around cross-domain and cross-infrastructure FAIR implementation through a case study driven methodology. Ultimately, it hopes to provide guidance for FAIR implementation both within specific domains and infrastructures and across them. The I and the R of FAIR pose particular challenges but are particularly important in addressing complex issues where datasets need to be combined and in enhancing scientific rigour and reproducibility. Consequently, increasing attention is being paid to semantics, the maintenance of referenceable vocabularies, terminologies and ontologies and to metadata profiles –and to tools that facilitate the tracking of provenance and process, or that use variable level metadata and semantics to facilitate data integration. The outputs will include recommendations for FAIR implementation for each case study, as well as a proposed Cross-Domain Interoperability Framework (CDIF). The CDIF Identifies a set of functional requirements for interoperability, particularly for steps in data combination, and recommends good practices for each of these requirements, in relation to the use of existing or emerging standards and specifications. The CDIF is categorically not a new standard, but is intended to act as a lingua franca across domain data practices and encourage the incorporation of a number of standards that perform important and specific functions across domains. We are keen to test this approach with colleagues from as many disciplines and application areas as possible. This talk will explore these developments in detail and make a case for the importance of the FAIR principles and related initiatives for the BIREDIAL community.

# CoARA: una coalición, otra iniciativa o la oportunidad definitiva para reformar la evaluación de la investigación

**EVA MÉNDEZ**

(ESPAÑA)

## RESUMEN EXTENDIDO

Méndez hace una invitación para participar activamente en COARA con el fin de promover acciones para la modificación de los criterios de evaluación académica. A 10 años de la creación de Research Data Alliance (RDA) aún no se vislumbran cambios en la forma de evaluar la ciencia, mientras se hace necesario definir una estructura que evalúe la ciencia tomando en cuenta el contexto. Algunas iniciativas buscan generar cambios en los sistemas de evaluación e incentivos que no conducen a la excelencia, sino a prácticas que afectan a la ciencia. Algunos ejemplos son ARRA (busca promover buenas prácticas de investigación), CAM (criterios para evaluar carrera académica), y Foro Latinoamericano sobre Evaluación Científica (FOLEC-CLACSO) (espacio de debate sobre evaluación). También se criticó los rankings, que han hecho tanto daño a las instituciones como el Factor de Impacto a los investigadores. Este tipo de mediciones generan un sistema científico de cantidad, no de calidad. Según Méndez, es necesario cambiar la forma en que se comunica, evalúa y se da crédito en la ciencia. Con el modelo actual el “crédito” de la calidad de la investigación se la está llevando el intermediario porque le traslada el “prestigio” a la revista como si publicar en ella tuviera alguna garantía de citación. Los más afectados son los investigadores jóvenes y las instituciones que producen menos. Con el actual acceso a internet, el antiguo modelo de publicación con intermediación no es necesario. La responsabilidad de medir la producción es de las instituciones, no de la industria.

# Infraestructuras, plataformas y tecnologías del Ecosistema Regional de Ciencia Abierta

**LAUTARO MATAS**

(ARGENTINA)

## RESUMEN

Red de repositorios de Ciencia Abierta constituida por una asociación de autoridades gubernamentales de Ciencia, Tecnología e Innovación, fundada el 29 de noviembre de 2012 en la Ciudad de Buenos Aires, mediante la firma de un Acuerdo de Cooperación, como resultado del proyecto financiado por el BID (2010-2013). LA Referencia se basa en una estrategia de red de nodos nacionales, cada país socio (nodo) agrega y aporta metadatos de insumos y resultados científicos en Ciencia Abierta a la red regional. Asimismo, las autoridades deben garantizar la sostenibilidad de sus nodos nacionales y de la red de acuerdo con los mecanismos jurídicos y presupuestarios que correspondan.

BIREDIAL  
ISTEC



Ponencias



# Alianza SIDALC: plataforma hemisférica que promueve el acceso abierto a información agropecuaria

## MANUEL HIDALGO OCONTRILLO

INSTITUTO INTERAMERICANO DE COOPERACIÓN PARA LA AGRICULTURA (IICA)

UNIDAD DE INFORMACIÓN Y DOCUMENTACIÓN DEL CENTRO DE SERVICIOS DE GESTIÓN DEL CONOCIMIENTO Y COOPERACIÓN HORIZONTAL

[manuel.hidalgo@iica.int](mailto:manuel.hidalgo@iica.int)

### EJE TEMÁTICO

Comunicación académica, científica y cultural en abierto

### RESUMEN

La Alianza de Servicios de Información Agrícola (SIDALC) es una alianza internacional de servicios de información agrícola, pecuaria, forestal y ambiental, cuyo éxito radica en el capital intelectual de más de 650 especialistas de 214 bibliotecas y 178 instituciones agropecuarias de 22 países que comparten información y servicios.

### PALABRAS CLAVE

Gestión del conocimiento; organizaciones comunitarias; agricultura; sistemas de información; herramientas de descubrimiento; repositorios digitales.

SIDALC, con el fin de contribuir al bienestar rural y mejorar la competitividad del sector agropecuario, busca poner a disposición desde un solo punto la información técnica y científica generada en América Latina y el Caribe (ALC), brindar apoyo a las instituciones del sector agropecuario, fortalecer alianzas con instituciones, redes y sistemas nacionales de información.

A partir de 2010 se inició el servicio de intercambio científico, que ha generado grandes beneficios para la comunidad de estudiantes e investigadores del sector con la distribución de más de 468.410 artículos especializados, lo que representa un total de 19,2 millones de dólares en recursos de conocimiento movilizados en la región.

Como parte de los procesos de modernización de la Alianza SIDALC, tomando en cuenta los aportes realizados en una evaluación realizada en el año 2021, dirigida a colaboradores, usuarios y equipo interno, se inició una transformación de la plataforma en un servicio que permita, a través de protocolos estandarizados, la captura automática de repositorios digitales de instituciones del sector agropecuario de ALC.

Entre los aspectos a implementar se encuentra la búsqueda general en todas las bases de datos participantes, búsqueda individual por base de datos, búsqueda por grupos de bases de datos pertenecientes a una institución, búsqueda a través de filtros de información preestablecidos, contenedores de centros de recursos especializados y recursos de conocimiento, plataforma de gestión de solicitudes de usuarios y bibliotecas participantes, gestión de perfiles profesionales e institucionales, migración de recursos de herramientas que no cuentan con estándares de intercambio, entre otros.

### **Conformación de la Alianza**

La Alianza de Servicios de Información Agropecuaria (SIDALC), es una alianza internacional de servicios información agrícola, pecuaria, forestal y ambiental, cuyo éxito radica en el capital intelectual generado por más de 650 especialistas de 215 bibliotecas y de 179 instituciones agropecuarias de 22 países que comparten información y servicios como una red. Creado en 1999, gracias al liderazgo del IICA y a los fondos recibidos por la Fundación Kellogg, hoy constituye una de las ventanas del conocimiento más importantes de América Latina y el Caribe (ALC).

SIDALC, con el fin de contribuir al bienestar rural y a mejorar la competitividad del sector agropecuario, busca hacer disponible desde un solo punto la información técnica y científica generada en ALC, brindar soporte a las instituciones de sector agropecuario, fortalecer alianzas con instituciones, redes y sistemas nacionales de información, promover la capacitación para el fortalecimiento de competencias de todos los miembros de la cadena del sector información, así como concientizar a los países sobre la importancia del uso responsable de la información y el conocimiento validado para contribuir con el crecimiento económico y social.

Desde sus inicios, SIDALC motivó el compartir el conocimiento que se encuentra en las instituciones agropecuarias (literatura gris). Ello permitió no solo complementar los servicios referenciales, los cuales facilitan de alguna forma la localización de información, sino también, iniciar con el manejo del texto completo (recursos digitales y electrónicos). A partir del año 2010, se inicia con el servicio de intercambio científico, lo que ha generado grandes beneficios a la comunidad de estudiantes e investigadores del sector con la distribución de más de 468.410 artículos especializados, lo que representa un total de US\$19,2 millones de dólares en recursos de conocimiento movilizados en la región.

Después de 23 años, se cuenta con la integración de 348 bases de datos que contienen 3.1 millones de recursos bibliográficos, de los cuales 301 671 corresponden a recursos digitales de libre acceso. En cuanto a los accesos (gracias a la indexación en Google y Google académico), se cuenta con el registro de 38.3 millones de usuarios de 208 países que han accedido y han dejado como resultado la visita a más de 223.4 millones de páginas. Los datos corresponden al corte de diciembre del 2021

Al realizar un análisis respecto al incremento de recursos digitales integrados a SIDALC a partir del año 2010 y hasta diciembre 2021, se registra un incremento promedio anual del 9 %. Si además se analiza esa misma variable, pero en bloques de cinco años, el incremento porcentual promedio es

del 25.6. Dicho análisis resulta interesante dado que refleja que, aunque existen picos positivos y negativos en algunos años, en promedio existe una tendencia de las instituciones a priorizar la creación y el acceso al recurso digital y electrónico.

En función de lo anterior, hoy en día, los recursos digitales constituyen el recurso más importante que los usuarios requieren, por lo que su demanda tiende a incrementar año con año, lo que hace indispensable trabajar en una estrategia para atender dicha demanda.

A través del tiempo la Alianza SIDALC ha motivado y recomendado a la comunidad de participantes, el uso de herramientas de código abierto desarrolladas por comunidades colaborativas, el uso de protocolos y estándares internacionales que promueva el intercambio de información, así como la digitalización de sus recursos documentales. Hoy en día, una gran cantidad de las instituciones participantes han migrado sus plataformas con el objetivo de exponer ante la comunidad internacional su producción técnica y científica.

A finales del 2020 y principios del 2021 se realizó una evaluación por parte de la Gerencia de Monitoreo y Evaluación del IICA, el informe evidencia los grandes beneficios tanto para los socios participantes como a los usuarios y consumidores de temáticas agropecuarias, también fue posible identificar varios puntos de mejora.

Por lo anteriormente expuesto, se plantea la posibilidad de modernizar el Descubridor Agropecuario de las Américas, que contemple varios aspectos clave para la difusión y acceso de recursos de información físicos y electrónicos/digitales, tomando como base las alianzas, que por más de 22 años el IICA ha promovido mediante la plataforma de la Alianza SIDALC.

## **Propuesta de modernización**

La propuesta para la conformación de un Descubridor Agropecuario para el hemisferio se basa en los servicios que brindan los colaboradores día con día a los participantes en Alianza SIDALC, estructura que es la clave de una red dinámica en la cual se comparte a la comunidad de usuarios para luego recibir mucho más, promoviendo así la democratización de la información y nuevos conceptos como ciencia abierta, la cual se fundamenta en producir conocimiento técnico y científico de manera abierta y colaborativa, además de hacer disponible todos los instrumentos utilizados de trabajo en el proceso.

A continuación, se detallarán los grandes módulos que serán transformados en el proceso de modernización iniciado: Plataforma tecnológica, Socios y colaboradores, Capacitación y asistencia técnica, Servicios de información.

### **1. Plataforma tecnológica**

#### **1.1 Recursos de información**

Como parte de los procesos de modernización de la Alianza SIDALC se propone convertir esta plataforma en un servicio que permita (mediante protocolos estandarizados) la captura automática de repositorios digitales de instituciones del sector agropecuario de ALC. La opción que se ha identificado lleva a la reconversión de la plataforma como un gran repositorio de recursos digitales distribuido y que, a su vez, también permita gestionar de forma individual cada base de datos compartida por los participantes.

Un valor agregado que permite esta opción es que facilita la creación de descubridores individualizados por institución, es decir, a quienes comparten varias bases de datos, se les habilita una opción para que puedan realizar búsquedas individualizadas en su grupo de bases de datos.

Los procesos de captura de información se han dividido en cinco contenedores principales, de los cuales, los últimos tres son nuevos recursos que se incorporan a la propuesta:

- a) Recursos bibliográficos digitales y referenciales
- b) Publicaciones periódicas (Revista)
- c) Cursos temáticos (Moodle)
- d) Datos de investigación
- e) Perfiles profesionales

Todos estos componentes tienen la característica que luego podrán ser cosechados por otras organizaciones a nivel internacional, permitiendo con esto una mayor exposición de la producción científica y técnica generada por instituciones de ALC.

## 1.2 Búsquedas de información

Uno de los principales componentes que, según los reportes estadísticos es el más utilizado, con un promedio de acceso de 3 millones de usuarios anualmente, es la plataforma de búsqueda de información, la razón radica en que desde un único punto se cuenta con acceso a recursos de información y especialistas de 215 bibliotecas de 179 instituciones agropecuarias de ALC. Este componente contempla las siguientes opciones de búsqueda, que además de fortalecer la plataforma, brinda opciones de gran importancia para los socios colaboradores que comparten sus recursos de información:

- a) Búsqueda general en todas las bases de datos participantes (Descubridor)
- b) Búsqueda individual por base de datos
- c) Búsqueda por grupos de bases de datos pertenecientes a una institución (Descubridores individualizados por institución)
- d) Búsqueda mediante filtros de información preestablecidos
- e) Contenedores de centros de recursos especializados para consultas temáticas

### **1.3 Diseminación e intercambio de recursos**

La diseminación es un aspecto que se analizó para fomentar una mayor exposición de los recursos de información, así como brindar una mayor visibilidad a las instituciones y especialistas de la región. Este tipo de procesos son posibles gracias a la gestión de protocolos automáticos y estándares internacionales para la gestión de la información. Entre los procesos mencionados se encuentran los siguientes:

- a)** Indexación automática en indexadores internacionales
- b)** Exposición de datos cosechados de las plataformas de nuestros socios colaboradores
- c)** Exportación en múltiples formatos para el beneficio de los usuarios
- d)** Exposición en redes sociales

### **1.4 Cosechador de noticias agropecuarias**

Un nuevo componente se integra como parte de la transformación de la Alianza SIDALC, el mismo permite, mediante protocolos de intercambio, capturar desde la fuente las noticias que comparten las instituciones relacionadas al sector agropecuario, en especial las instituciones colaboradoras de la Alianza.

## **2. Socios y colaboradores**

Los principales socios y colaboradores, proveedores de recursos de información, a lo largo del tiempo se han identificado en las siguientes instancias:

### **2.1 Redes nacionales de información agropecuaria**

El principal contacto y acceso a instituciones agropecuarias de los países de ALC son las redes nacionales de información, mismas que se han transformado para proveer un mejor servicio a sus asociados, la propuesta pretende tener un

mayor acercamiento con el fin de promover acciones de modernización de sus colaboradores y de intercambio abierto de recursos de información.

## **2.2 Instituciones agropecuarias de ALC**

Las instituciones que no se organizan en redes nacionales de información también pueden formar parte de la Alianza SIDALC, ellas también se han fortalecido con los servicios brindados en la red. Las mismas cuentan con bibliotecas especializadas, que registran en sus plataformas, importantes recursos técnicos y científicos al servicio de la comunidad internacional.

## **2.3 Bibliotecas especializadas**

Se han identificado una importante cantidad de bibliotecas especializadas en la región, mismas que cuentan con un perfil en la base de datos del directorio de bibliotecas. Este recurso será transformado con el objetivo de permitir a las instituciones la autogestión de sus perfiles, así como su localización geográfica.

## **2.4 Comunidades de especialistas**

Esta comunidad de especialistas es una pieza fundamental en los procesos de atención de solicitudes que ha dado importantes aportes a la red, sin embargo la característica más destacada como aporte es su gran disposición en compartir y colaborar con los usuarios en la búsqueda de información.

## **3. Capacitación y asistencia técnica**

Este componente está muy relacionado con los protocolos de registro y actualización de información. Nuestros colaboradores y proveedores de información requieren avanzar en sus procesos de modernización por lo que este espacio pretende, por un lado, motivar su transformación y por otro lado acompañar con la puesta en marcha de acciones clave para su modernización, tanto con herramientas, como con esquemas de metadatos aptos para su integración en la plataforma de la Alianza SIDALC.



### **3.1 Alfabetización informacional**

La alfabetización informacional se enfoca en dos grupos principales, el primer grupo, son los socios colaboradores de la red a los que se les motiva a transformar sus plataformas y servicios. El segundo grupo, son los usuarios en general que requieren de algún tipo de asistencia para que puedan realizar consultas eficientes de los datos requeridos.

### **3.2 Plataformas de gestión de información**

Las plataformas de gestión de información son las herramientas más importantes para los procesos de registro y consulta de recursos documentales de las instituciones. La implementación y el uso de herramientas modernas, que cumplan con estándares internacionales en cuanto a esquemas de metadatos y protocolos de intercambio de información, permiten que la cosecha de datos sea más eficiente, favoreciendo así los tiempos de respuestas en su actualización y disponibilidad. Como proveedores de recursos de información se hace necesarios que nuestros socios cuenten con recursos aptos que faciliten los procesos de captura.

### **3.3 Perfiles profesionales**

Este es uno de los componentes que se incorporan directamente a la plataforma, puesto que constituye una herramienta que promueve la visualización y exposición de los especialistas de las instituciones participantes, por lo que se hace necesario acompañar a las instituciones en los procesos de ingreso y registro que esta implica.

## **4. Servicios de información**

Conocido como Servicio de Intercambio Científico, este componente es parte de la Alianza SIDALC y en este momento se transforma con una nueva herramienta que hace más eficiente la atención de solicitudes tanto a nivel interno de la red como a lo externo. El recurso contempla un espacio más

dinámico y simple que facilita la atención y registros de cada una de las acciones realizadas, procurando así una efectiva y eficiente respuesta.

Se contempla la participación de entidades gubernamentales, instituciones nacionales e internacionales relacionadas al sector agropecuario del ALC, que promuevan el intercambio de recursos de información para el beneficio del sector.

El análisis estadístico será parte importante de este componente, debido a que las herramientas permitirán registrar varios datos que serán clave para la toma de decisiones relacionadas con aspectos técnicos de la Alianza, pero de importancia para las instituciones dado que se contará con mejor información en cuanto al uso y requerimientos de nuestros usuarios. Se contemplan los siguientes componentes:

#### **4.1 Miembros de la red de la Alianza SIDALC**

Los miembros de la red cuentan con varios servicios en sus bibliotecas y unidades de información, mismos que ponen a disposición de la red. Los recursos que no se localizan en la plataforma de la Alianza SIDALC son solicitados a la comunidad la cual comparte abiertamente la información requerida.

#### **4.2 Usuarios**

Los usuarios tienen la posibilidad de acceder los recursos digitales por medio de la plataforma, pero conscientes de que aún existen muchos recursos en formato físico, se abre la posibilidad de que mediante un formulario se realicen las solicitudes directas a las bibliotecas responsables de la información, donde cada biblioteca toma contacto con el usuario para hacerle llegar la información requerida, siempre y cuando cumpla con los derechos de propiedad intelectual.

### **4.3 Chat**

El chat es una herramienta que se incorpora con el objetivo de acercar a los especialistas de información con los usuarios, con el objetivo de brindar una atención más directa y personalizada para atender consultas puntuales en tiempo y forma.

### **4.4 Comunidades de práctica (virtuales)**

Este nuevo recurso está muy relacionado con los Centros de Recursos de Información (CRI) en el que se filtran contenidos sobre temáticas específicas, por ejemplo: Agricultura orgánica, Roya del café, Cambio climático, Mujer rural, entre muchos otros. Básicamente se contempla incorporar a los CRI un grupo de herramientas, experiencias, redes y otros componentes relacionados con la temática desarrollada. El objetivo principal es contar con un espacio temático en el cual se suman recursos y se fomenta el intercambio entre los interesados.

### **4.5 Análisis de datos estadísticos**

El recurso contempla el seguimiento estadístico dinámico del uso de las plataformas de gestión de información y atención de solicitudes, ambos recursos también exponen las necesidades de información de los usuarios, por consiguiente, será un importante insumo para visualizar las necesidades reales en función de las cifras reportadas.

Actualmente el proyecto se encuentra en desarrollo en su primera etapa. Debido a aspectos presupuestarios se contempla la incorporación de la cosecha de cursos temáticos, análisis de datos, comunidades de prácticas y el cosechador de noticias agropecuarias para una segunda etapa.

## **Sostenibilidad**

En el estudio y reporte realizado a inicios del 2021 se comparten algunas ideas para la sostenibilidad de la Alianza SIDALC, por lo que se sugieren las siguientes acciones:

- 1) Los proyectos que emprenda el IICA con otras organizaciones del sector, se sugiere que sus bibliotecas formen parte de la Alianza, con el objetivo de poner al servicio de la comunidad internacional todo el material de su institución, por consiguiente, tendrán acceso a una importante red de información agropecuaria de ALC.
- 2) Se sugiere que los nuevos proyectos que implemente el IICA, asignen un porcentaje anual para la sostenibilidad de la Alianza, plataformas y servicios.
- 3) Membresías con aportes anuales.
- 4) Donaciones.

## Bibliografía

- ANDRADE, M. (2021). Evaluación Prospectiva de la Alianza de Servicios de Información Agropecuaria de las Américas (SIDALC).
- GRANADOS, C. (2004). Sistema de información y documentación agropecuaria de América (SIDALC). Evaluación Final.
- HIDALGO, M. (2022). AgriPerfiles: plataforma hemisférica que promueve una mejor vinculación entre profesionales de la agricultura y el desarrollo rural. En: V Congreso Internacional y V Encuentro Nacional de Extensión Rural. <<https://agriperfiles.agri-d.net/display/n35877>>
- HIDALGO, M. (2022). Alianza SIDALC: Innovación focalizada en productos y servicios para el sector agropecuario. En 48th IAMSLIC Conference & 4th Grupo Regional Latinoamericano. <<https://agriperfiles.agri-d.net/display/n455834>>

# Automatización de la evaluación de repositorios institucionales

## **JUAN SEBASTIÁN GONZÁLEZ SANABRIA**

UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA

[juansebastian.gonzalez@uptc.edu.co](mailto:juansebastian.gonzalez@uptc.edu.co)

## **FABIÁN NICOLÁS RAMOS CORREDOR**

UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA

[fabian.ramos@uptc.edu.co](mailto:fabian.ramos@uptc.edu.co)

## **GERMÁN AMÉZQUITA BECERRA**

UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA

[german.amezquita01@uptc.edu.co](mailto:german.amezquita01@uptc.edu.co)

### **EJE TEMÁTICO**

Evaluación y métricas alternativas

### **RESUMEN**

El auge de los repositorios digitales ha enmarcado un avance significativo en el acceso al conocimiento académico y científico, aumentando su impacto debido a su mayor alcance y menor costo. Sin embargo, estas plataformas son un tema novedoso que, en un principio, no contó con estándares o modelos para su implementación y funcionamiento. Es por ello que se presentaron inconsistencias entre repositorios en temas como interoperabilidad y preservación digital, entre otros. A lo largo del tiempo han aparecido diversas guías para la evaluación e implementación de repositorios, tales como la de RECOLECTA o el certificado DINI (Deutsche Initiative für Netzwerkinformation). Estas últimas pretenden evaluar las plataformas en su totalidad e incluyendo diferentes elementos. Sin embargo, no lo hacen de manera automática, lo que hace subjetiva su fiabilidad. Por lo anterior, se desarrolló un aplicativo para la automatización de la evaluación de repositorios que, de manera sistematizada, examina la mayoría de las características encontradas en las guías de evaluación, especialmente en lo referente a interoperabilidad, generación de contenidos y uso de estándares internacionales para la gestión de

información. De allí se obtiene una herramienta que coadyuva a las instituciones que lo utilicen a conocer las características y funcionalidades por mejorar en sus repositorios institucionales.

#### **PALABRAS CLAVE**

Automatización de proceso; desarrollo de software; evaluación; repositorios institucionales.

### **Introducción**

Un repositorio se define como “un sitio web que recoge, preserva y difunde la producción académica de una institución (o de una disciplina científica), permitiendo el acceso a los objetos digitales que contiene y a sus metadatos” (González et al., 2015, citando a López, 2013). Se enmarcan cuatro características que deberían poseer los repositorios institucionales: el autoarchivo, que implica que el creador, propietario o un tercero deposite el contenido en la plataforma; la interoperabilidad, que comprende el uso de procesos normalizados que permitirán la comunicación con otros repositorios a través de OAIPMH (Open Archives Initiative – Protocol Metadata Harvesting); el acceso libre y gratuito al texto completo; y la preservación a largo plazo (González et al., 2015). Este último, especialmente en el marco del movimiento de acceso abierto y ciencia abierta.

Hoy día, los repositorios institucionales comprenden uno de los medios o plataformas virtuales para la divulgación del conocimiento con mayor aumento con respecto a otras opciones como las revistas. Esto se le puede atribuir a las diferentes iniciativas sobre acceso abierto que se han presentado en el último siglo, como es el caso de la denominada Budapest Open Access Initiative (2002), una de las más importantes. Esta iniciativa, también llamada BOAI, invita a sus firmantes a promover y fomentar el Internet como una herramienta por medio de la cual se exponga y comparta la información académica y científica de forma gratuita.

Es por lo anterior que los repositorios digitales se han convertido en plataformas que ganaron importancia a causa del movimiento denominado Acceso Abierto, pues representaban el medio idóneo para la divulgación del conocimiento. Esto se ve reflejado en el aumento exponencial de repositorios a nivel mundial, según OpenDOAR (Directory of Open Access Repositories) (JISC, s.f.). En agosto de 2006, había 501 repositorios digitales registrados, y había 5778 en noviembre de 2021.

Sin embargo, debido a la falta de una normalización de los repositorios en cuanto a cómo deberían funcionar y qué deberían contener, su implementación se llevaba a cabo sin seguir lineamientos o estándares, lo que desencadenó en un problema para los usuarios de estos sistemas. A raíz de lo anterior, diferentes organizaciones, instituciones e investigadores se han puesto a la tarea de establecer modelos o guías con aspectos que permitan evaluar la calidad de los repositorios. Dichos criterios pueden estar enfocados en la evaluación de algunos temas en particular (como la usabilidad o el diseño del repositorio) o abarcar también una visión más general, donde se incluyen elementos como políticas e interoperabilidad, entre otros.

Como una de las primeras iniciativas, el grupo de trabajo de la Deutsche Initiative für Netzwerkinformation (DINI, Iniciativa Alemana para la Información en Red) (AMH et al, s.f.) definió criterios y formalizó el Certificado DINI para Servicios de Documentación y Publicación (Grupo de trabajo “Publicación Electrónica”, 2012). La certificación se compone de los criterios: visibilidad del servicio, políticas, asesoramiento a autores y editoriales, aspectos legales, seguridad de la información, indexación e interfaces, estadísticas de acceso y disponibilidad a largo plazo.

Las recomendaciones, pese a que no son obligatorias, deben tenerse en cuenta, pues esta certificación está en una continua actualización, y puede que, en una nueva versión, alguna de las recomendaciones se convierta en requisito mínimo. La DINI, al igual que el certificado, surgió principalmente para

contribuir en la mejora de las plataformas universitarias de divulgación de conocimiento en Alemania, por lo cual no es tan conocida fuera de esta región.

En el mismo sentido, el proyecto DRIVER (Digital Repository Infrastructure Vision for European Research), que se unió a OpenAIRE (Open Access Infrastructure for Research in Europe), define directrices enfocadas en la comunicación entre repositorios, buscando establecer un estándar de compatibilidad e interoperabilidad que permita integrar los repositorios con otras plataformas, como es el caso de los recolectores. Los criterios se pueden dividir en dos capas: la sintáctica (utilización de OAI-PMH y de OAI\_DC) y la semántica (uso de vocabularios) (Vanderfeesten et al., 2008). Sin embargo, la guía propuesta comprende las siguientes secciones: a) recursos textuales, b) metadatos (Dublin Core) y c) implementación de OAI-PMH.

De la misma forma que la propuesta de la DINI, por cada sección de este modelo existen unas características o puntos obligatorios y otros recomendables. En resumen, las directrices presentadas por DRIVER pretenden establecer un léxico común y una base para los datos de OAI-PMH y Dublin Core, con el fin de garantizar la interoperabilidad entre repositorios.

OpenAIRE publicó en 2010 las Directrices para administradores de repositorios de literatura (OpenAIRE, 2020), las cuales están basadas en las directrices DRIVER y brindan orientación a los administradores de repositorios digitales para que definan e implementen políticas de gestión de datos. Esta guía describe cada uno de los metadatos que deben poseer los documentos de los repositorios, estableciendo una definición, una obligatoriedad y unas recomendaciones para cada elemento. Además, la última versión (v4 – 2018) incorpora esquemas de metadatos distintos como DC (Dublin Core) y DataCite (DataCite Metadata Working Group, 2021).

Estos esquemas de metadatos han sido implementados, pues, al no existir un estándar internacional, los repositorios implementan un esquema u otro dependiendo sus necesidades o requerimientos. Las directrices de OpenAIRE



contienen un total de 32 atributos, lo cual implica que no se puede cumplir con lo sugerido en algunos esquemas, como es el caso de Dublin Core, que solo establece 15 atributos, aunque, para poder dar cumplimiento a la totalidad de la guía, se podrían complementar los elementos de un esquema con el de otro, o directamente implementar un esquema más completo con respecto a lo establecido por OpenAIRE.

Ahora bien, el proyecto RECOLECTA (o Recolector de Ciencia Abierta), surgido en el 2007 como una colaboración entre la Fundación Española para la Ciencia y la Tecnología (FECYT) y la Red de Bibliotecas Universitarias (REBIUN), creó la Guía para la evaluación de repositorios institucionales de investigación (Azorín et al., 2021), que se define como una herramienta para la autoevaluación de repositorios. La guía se compone de los siguientes criterios:

- Visibilidad
- Políticas
- Aspectos legales
- Metadatos descriptivos de la publicación
- Interoperabilidad de metadatos y acceso a los contenidos
- Logs y estadísticas
- Seguridad, autenticidad e integridad de los datos
- Servicios y funcionalidades de valor añadido

Esta guía, que se encuentra en su cuarta edición, incluye novedades como la adición de vocabularios para determinados metadatos elaborados por COAR (Confederation of Open Access Repositories), el esquema de metadatos DataCite y niveles definidos por la National Digital Stewardship Alliances (NDSA) para el tema de preservación digital de contenidos. Además, para la validación de los metadatos, se utilizan como referencia las directrices establecidas por OpenAIRE. Consecuentemente con el respaldo de las instituciones que la proponen, y con la inclusión de otras guías y estándares

reconocidos, esta propuesta de evaluación es una de las más conocidas y aceptadas por la comunidad.

Por último, pese a que en los últimos años se han definido modelos de evaluación más completos y reconocidos, el software que existe para realizar dichas evaluaciones es escaso. En Latinoamérica, un grupo interdisciplinar desarrolló un aplicativo web de software libre llamado dPyx – Herramienta de autoevaluación de sistemas de información académica y científica (Escire, 2019), con el objetivo de dar a los administradores de plataformas de información un conjunto de indicadores que permitieran evaluar sus sistemas.

Los indicadores de dPyx están basados en buenas prácticas, documentación y estándares internacionales como ISO 16363:2012, OpenAIRE, WCAG y OAIS. En esta plataforma se definieron ocho criterios o secciones (Escire, 2019), a saber:

- Gobernanza: mandatos, políticas, recursos, financiamiento
- Mantenimiento y desarrollo: lineamientos, colecciones, roles, procesos
- Accesibilidad: plataformas, velocidad, formatos, permanencia
- Software: estabilidad, actualización, protocolos, seguridad
- Hardware: actualización, mantenimiento, conectividad
- Preservación digital: formatos, licencias, procesos, estándares
- Posicionamiento y visibilidad: índices, buscadores, directorios, metadatos, interoperabilidad
- Ética e integridad: buenas prácticas, transparencia, FAIR

En esta plataforma se manejan tres roles: usuarios (repositorio/revista), administrador y evaluador. También se cuenta con modelos de evaluación para revistas científicas y repositorios digitales. La operación del sistema comienza con la solicitud para la creación de un administrador que establecerá los evaluadores y usuarios. Luego, los usuarios son aquellas personas que representan una revista o repositorio y se encargarán de realizar la evaluación

por medio de un cuestionario. Y en última instancia se encuentran los evaluadores, cuya función es validar que la información ingresada por los usuarios sea correcta, además de dar observaciones y aprobar o no la finalización de los criterios de calificación (Escire, 2019).

Por lo anterior, esta investigación presenta el desarrollo de un aplicativo para la automatización de la evaluación de repositorios, el cual se ejecuta en tres partes: la primera consiste en la definición de los criterios de la evaluación; la segunda corresponde al establecimiento de herramientas y servicios para realizar la automatización; y la última es el desarrollo del software. Lo anterior, con el objetivo principal de brindar a la comunidad una herramienta que coadyuve en el proceso de mejora del acceso abierto en la ciencia.

### **Metodología**

El desarrollo del presente trabajo comprendió tres partes. La primera correspondió al método de revisión sistematizada de literatura, en el cual se buscaron e identificaron diferentes propuestas para la evaluación de repositorios en los recolectores Google Scholar, el Recolector de Ciencia Abierta (RECOLECTA) y LA Referencia. Luego, se descartaron las propuestas que ya se contemplaban en guías más actuales o que involucraban aspectos no tan relacionados con los repositorios como plataformas, como el caso de La accesibilidad web en los repositorios institucionales – la UOC a examen (Bolaños, 2012), en la cual se evalúa el repositorio con una norma referente a diseño accesible que involucra a cualquier tipo de plataforma web. A partir de la información filtrada, se consolidó una guía de evaluación conformada por los diferentes criterios planteados en otras guías.

En la segunda parte se determinó la viabilidad de la automatización de cada uno de los elementos de la guía, teniendo en cuenta diferentes aspectos como el nivel de acceso (pública o privada) requerido para obtener la información o la localización de los datos si se trata de algo normalizado dentro o fuera del

repositorio. Posteriormente, se identificaron las herramientas y servicios externos que permitieron obtener la información que dio respuesta a los ítems que se automatizaron.

En la tercera y última parte se desarrolló el prototipo de evaluación a partir de la guía planteada en la primera sección, así como las herramientas y servicios definidos en el segundo apartado. Este prototipo se realizó haciendo uso del marco de trabajo Scrum, en el cual se plantearon 4 sprints o iteraciones y, al finalizar cada una, se realizó la presentación y revisión de los criterios desarrollados para establecer si la implementación fue adecuada conforme a los resultados obtenidos. Además, posterior a la creación del prototipo, se realizó un proceso de validación y verificación general de los resultados. Esto último, en pro de revisar la veracidad de la información provista en los elementos automatizados, además de validar si alguno de los elementos automatizados presentaba alguna falla, dado que en estos casos se depende de un servicio o página web externa.

## **Resultados**

En la revisión documental realizada se encontraron tres guías con amplio reconocimiento internacional: RECOLECTA (Azorín et al., 2021), DINI (Grupo de trabajo "Publicación Electrónica", 2012) y ALICIA (RNRPCTIAA, 2020). Esta última se basa en las dos primeras (principalmente en la de RECOLECTA), pero adiciona definiciones en la parte de metadatos y aspectos legales, además de incluir un apartado denominado soporte informático, enfocado en la implementación y mantenimiento de los repositorios.

Por otra parte, las guías de RECOLECTA y de DINI son similares en la mayor parte de sus criterios, solo variando la cantidad de elementos que poseen cada uno de ellos. En este aspecto, RECOLECTA tiene un mayor número de elementos, lo que representa un mayor detalle en la evaluación; por ejemplo, en el apartado de metadatos de la guía de DINI solo se revisan elementos y

características básicas del esquema DC, mientras que RECOLECTA tiene en cuenta 3 esquemas distintos y cuenta con un vocabulario para algunos de los elementos.

Teniendo en cuenta lo mencionado anteriormente, para la definición de la guía a automatizar, se tomó como referencia la de RECOLECTA. La guía de evaluación definida comprende 8 criterios:

- Visibilidad. En este apartado se mencionan los aspectos que dan al repositorio un mayor reconocimiento, en una búsqueda por publicitar la plataforma y su contenido.
  - Presencia en directorios internacionales: OpenDOAR, ROAR, OAI Data Providers, re3data
  - Presencia en recolectores internacionales: LA Referencia, OpenAIRE, Google Académico, CORE, BASE
  - Presencia en recolectores nacionales
  - Existencia de nombre normalizado del RI en directorios y recolectores
  - Existencia de URL segura (https) y amigable (nombre del RI)
  - Disponibilidad de documentos en acceso abierto
  - Existencia de iniciativas para fomentar la visibilidad del repositorio dentro de la propia institución
- Políticas. Este apartado se refiere a temas de organización y gobernanza del repositorio, en aras de conocer su estado y avance en cuanto a la definición de directrices, normativas, actividades y procesos.
  - Existencia de una política institucional de acceso abierto
  - Adhesión a la declaración de Budapest, uno de los documentos fundacionales del movimiento de acceso abierto
  - Existencia de una política de actuación del RI (documento público unificado)

- Existencia de información de la política de forma dispersa en el sitio del RI
- Indicación de misión y objetivos del RI
- Indicación de quién puede depositar, qué se puede depositar y en qué formatos
- Indicación de cómo se lleva a cabo la preservación de los contenidos
- Indicación acerca de la reutilización de los metadatos
- Existencia de datos de contacto y/o asesoramiento visible
- Aspectos legales. Este apartado describe el manejo de los derechos de autor.
  - Exigencia al autor de reconocer que no está infringiendo ningún derecho de propiedad intelectual
  - Exigencia al autor de la firma de una autorización para la distribución de su obra
  - Mención de cómo puede hacer el autor para saber si su obra es depositable según la política editorial (Sherpa/Romeo, Dulcinea)
  - Inclusión de los derechos de autor en los metadatos de cada recurso
  - Inclusión de los derechos de autor en cada recurso
- Metadatos. Los metadatos son información estructurada o semiestructurada que describe el contenido, la calidad, las condiciones, la historia, la disponibilidad y otras características de los documentos, incluyendo datos como autores, fecha de publicación, referencias, lenguaje y tipo, entre otros. Esta sección define las características, formato y vocabulario de los metadatos que debe contener cada documento del repositorio.
  - Uso del esquema de metadatos Dublin Core (DC)
  - Inclusión de identificadores de autor (ORCID, IraLIS)
  - Inclusión de los siguientes campos:

- Autor (dc:creator)
- Título (dc:title)
- Tipo de resultado de investigación (dc:type)
- Versión del recurso (dc:type)
- Fecha de publicación (dc:date)
- Derechos de autor (dc:rights)
- Descripción (dc:description)
- Formato (dc:format)
- Idioma (dc:language)
- Identificador (dc:identifier)
- Temática/descriptores/palabras clave (dc:subject)
- Colaboración (dc:contributor)
- Referencia de financiación (dc:relation)
- Editorial (dc:publisher)
- El campo de derechos de acceso debe diligenciarse conforme al vocabulario establecido (closedAccess, embargoedAccess, openAccess, restrictedAccess)
- El campo de fecha de publicación debe diligenciarse conforme al formato establecido (ISO 8601 – AAAA-MM-DD, AAAA-MM-DDTHH:MM:SSZ)
- El campo de idioma debe diligenciarse conforme al vocabulario establecido (ISO 639-1, 639-2 y 639-3, código zxx)
- El campo tipo de resultado de investigación debe contener una única ocurrencia
- El campo tipo de resultado de investigación debe asignarse según el vocabulario de tipos de recursos de COAR (Anexo1)
- El campo de formato debe asignarse conforme al vocabulario establecido (Anexo2)
- El campo de versión del recurso debe contener una única ocurrencia

- El campo de versión del recurso debe diligenciarse conforme al vocabulario COAR (draft, submittedVersion, acceptedVersion, publishedVersion, updatedVersion)
- Debe aplicarse algún sistema de clasificación normalizado (disposición de uno o varios sistemas de clasificación normalizados tales como CDU, JEL, UNESCO)
- Debe utilizarse algún esquema de metadatos técnicos y/o de preservación
- El repositorio debe desarrollar alguna actividad de curación de metadatos
- Deben exportarse metadatos a otro formato distinto de DC
- Interoperabilidad. Este apartado atañe a la declaración de procesos y características en los servicios de extracción de contenido de la plataforma.
  - Es recolectado por LA Referencia-OpenAIRE
  - Se proveen los metadatos a través del protocolo OAI-PMH
  - Se marcan los registros eliminados
  - El tiempo de vida del testigo de reanudación es de un mínimo de veinticuatro horas
  - El correo electrónico del administrador del repositorio está disponible en la etiqueta AdminEmail dentro de la respuesta a una orden Identify
  - Existe una declaración de Description en la respuesta a una orden Identify
  - La entrega de registros a través del protocolo OAI-PMH es progresiva y a través de lotes
  - El tamaño de los lotes para la entrega de registros está dentro del rango de 100-500 registros
  - El formato de la fecha expresado en la orden Identify coincide con el campo datestamp de los registros
  - Se integra con otros sistemas de información de la institución



- Incluye etiquetas <meta...> en las **cabeceras HTML**
- El repositorio soporta otros protocolos y APIs para compartir metadatos y/o contenidos
- Hace un uso extendido de identificadores persistentes (DOI, Handle, URN, ORCID)
- Seguridad. Este apartado corresponde a la evaluación de las prácticas y estrategias usadas por los administradores de los repositorios para mantener la integridad y la fiabilidad de la información, tanto en donde se encuentra almacenada como en los procesos en los cuales se transfiere.
  - Se hace mención de la realización de copias de seguridad en el sitio del RI
  - Se hace mención de la ejecución de sumas de verificación (checksum) en el sitio del RI
  - Existen, como mínimo, tres copias de los registros (metadatos y ficheros), y por lo menos una de ellas está ubicada en una localización geográfica distinta
  - Se realiza identificación, control y validación de formatos (JHOVE, DROID, Xena)
- Estadísticas y logs. Este criterio revisa la información sobre el acceso y uso de la plataforma por parte de los usuarios, tanto a nivel general como por cada documento. Además, se verifica la forma como se almacena esta información.
  - Hay disponibilidad de estadísticas públicas del RI en general
  - Hay disponibilidad de estadísticas públicas de cada documento depositado en el RI
  - Los logs del servidor web donde está alojado el repositorio se archivan de forma permanente
  - Se utiliza el estándar COUNTER

- Servicios de valor añadido. Las características novedosas, diferentes y que aportan valor son un tema primordial para mejorar la percepción por parte de los usuarios. Por tanto, en este ítem se evalúa si el repositorio posee servicios diferenciales que lo pueden posicionar frente a otras plataformas, tales como:
  - Uso de redes sociales para compartir cada documento (twitter, Facebook, LinkedIn)
  - Integración con gestores bibliográficos (Zotero, Mendeley)
  - Visualización/exportación de los metadatos en diferentes esquemas (METS, PREMIS, RDF, JSON, MARC, BibTeX)
  - Disponibilidad de servicios de alerta (RSS)
  - Existencia de perfiles de autor
  - Métricas basadas en citas
  - Métricas de nueva generación (como el índice H)
  - Herramientas y servicios externos

Luego de haber establecido los elementos de la guía, se determinó que automatizar la totalidad de estos era inviable, pues la información en algunos ítems se obtenía de webs externas al repositorio o no estaba publicada en Internet. El primer caso presenta el inconveniente de que las páginas externas eventualmente presentan una organización y estructura diferente, por lo cual la información no siempre se va a encontrar en el mismo sitio. Asimismo, en la segunda opción, los datos requeridos los conocían únicamente los administradores del sistema.

Algunos ejemplos de estos casos presentados son la indicación de quién puede depositar, qué se puede depositar y en qué formatos (políticas), donde la información se encuentra en la red, pero su ubicación no está normalizada. También está el caso del ítem los logs del servidor web donde está alojado el repositorio se archivan de forma permanente, en el cual los datos son propios del funcionamiento del aplicativo.

Dejando de lado los ítems de los casos que no se iban a automatizar, se procedió a determinar las herramientas y servicios para los demás elementos. Primero, se identificó que, para los ítems relacionados con OpenDOAR y Google Scholar, se podía hacer uso de APIs propias o externas de estas plataformas, específicamente Sherpa API y Scale SERP para la primera y segunda respectivamente. Para los demás recolectores y directorios, se usó la técnica denominada web scraping, dado que la información requerida estaba presente en el HTML de las diferentes páginas web de las plataformas.

Para los directorios ROAR y OAI Data Providers, se realizó un proceso distinto, debido a que el primero no permitía hacer consultas por medio de web scraping, y el segundo tenía una página estática que no se actualiza frecuentemente, lo cual hacía ineficiente descargar la página en cada evaluación, pues se obtenía la misma información la mayor parte del tiempo. Para solucionar esto, se creó un script que descarga los datos de OAI Data Providers y carga un archivo XML, que se obtiene en la página web de ROAR, para luego crear una base de datos relacional con la información de estas dos plataformas.

Por último, para todo lo referente a los metadatos y atributos de los documentos o trabajos publicados en los repositorios, se usó web scraping, buscando, en el documento HTML, las etiquetas correspondientes a los elementos de interés, como por ejemplo <meta> para los datos del esquema DC. Sin embargo, las URL de los documentos dependen de los hallazgos que se hagan en los directorios y recolectores, pues cada repositorio organiza la información según sus necesidades y disciplinas. Es por esto que no era viable estandarizar la obtención de los links directamente de la búsqueda en los repositorios. Esta dependencia implica que, si no se encuentra información en la primera parte, no se evaluarán los ítems relacionados con los documentos.

Para el web scraping, se usó específicamente el método llamado Document Object Model (DOM) Parsing, en el cual se convierte el HTML de la página web en un árbol DOM, del cual se pueden obtener los datos por medio de la

búsqueda por sus atributos, etiquetas o relaciones. Un ejemplo que representa lo anterior se muestra en la Figura 1.

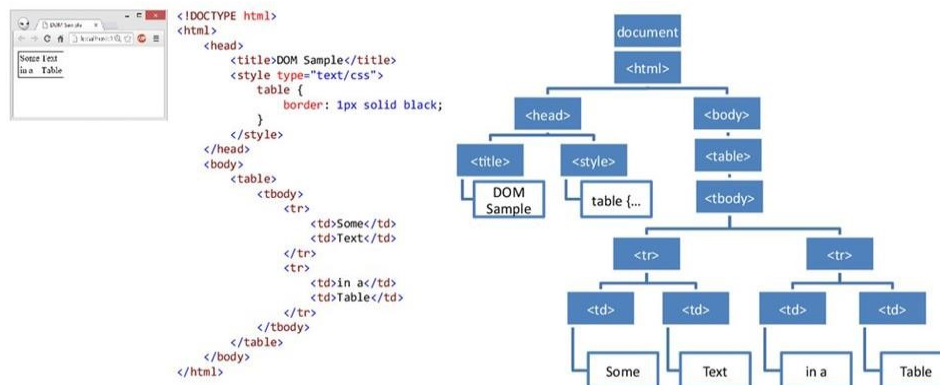


FIGURA 1. Ejemplo de Document Object Model (DOM) Parsing, extraído de la presentación Module 7: Accessing DOM with JavaScript (SoftServe, 2019)

Para el desarrollo del aplicativo, se definió una estructura cliente-servidor, con el fin de que se pudieran tener diferentes usuarios en simultáneo. Posterior a esto, se eligieron dos frameworks para el software: React para la parte del cliente y Flask para el servidor. Estos frameworks fueron seleccionados por el conocimiento propio de los investigadores, además del hecho de que Flask, al estar basado en Python, ofrecía una ventaja en lo relacionado a web scraping – este lenguaje es uno de los que tiene las librerías más completas para el desarrollo de esta técnica. En este apartado se describen las pantallas del sistema, junto con sus respectivas funcionalidades.

En la página principal del aplicativo (<https://criteria.onrender.com/>) se encuentran dos componentes principales: el primero es un formulario para comenzar la evaluación del repositorio, y la segunda es un menú con las opciones de Comenzar evaluación y Evaluaciones existentes (Figura 2).

**Evaluar repositorio**

Enlace al repositorio\*  
http:// - https://

Nombre del repositorio\*  
Nombre

Nombre alternativo  
Nombre del repositorio (opcional)

**Evaluar**

FIGURA 2. Página de inicio

Los datos requeridos en el formulario son los siguientes: enlace al repositorio, nombre del repositorio y nombre alternativo (opcional). El enlace se solicita debido a que, en ciertos ítems, se valida que la URL ingresada coincida con la del repositorio; y los dos nombres se utilizan para realizar las búsquedas en los recolectores y directorios internacionales de donde se obtienen las URL de los documentos a evaluar. Además, cabe aclarar que se solicitan dos nombres por la falta de normalización de estos aspectos en las diferentes plataformas.

Posteriormente, para cada característica se solicitan una serie de datos. Por ejemplo, en Visibilidad se encuentra un menú con los diferentes criterios y unas tarjetas con los diferentes ítems correspondientes a este aspecto. En los elementos que son automáticos, se encuentra una etiqueta que lo menciona y, en los que no, está la opción de definir si se cumple o no la condición (Figura 3).

**Visibilidad** Total 0

1. Presencia en directorios internacionales: OpenDOAR, ROAR, OAI Data Providers, re3data Automático

2. Presencia en recolectores internacionales: LA Referencia, OpenAIRE, Google Académico, CORE, BASE Automático

3. Presencia en recolectores nacionales

Sí  No

4. Existencia de nombre normalizado del RI en directorios y recolectores Automático

5. Existencia de URL segura (https) y amigable (nombre del RI) Automático

6. Disponibilidad de documentos en acceso abierto Automático

7. Existencia de iniciativas para fomentar la visibilidad del repositorio dentro de la propia institución

Sí  No

Guardar

FIGURA 3. Página de visibilidad

Para los recolectores y directorios internacionales, se verifica que alguno de los nombres encontrados en estas plataformas tenga una similitud de por lo menos 90 % con respecto a los ingresados en la página de inicio. Esto, dado que, en algunos casos, los nombres no coincidan en su totalidad por acentos o conectores. En el ítem de presencia en recolectores nacionales, si es afirmativa la respuesta, se debe ingresar mínimo una URL de las cinco entradas posibles, como se muestra en la Figura 4.

3. Presencia en recolectores nacionales

Sí  No

Enlace de recolector\*  
http// - https//

Enlace de recolector  
http// - https//

Enlace de recolector  
http// - https//

Enlace de recolector  
http// - https//

Enlace de recolector  
http// - https//

FIGURA 4. Ítem de presencia en recolectores nacionales

La evaluación del cuarto elemento consiste en verificar si los nombres encontrados en los directorios y recolectores son iguales entre sí. El quinto ítem cuenta con dos partes: la primera verifica que la URL del repositorio contenga https, y la segunda que la misma no tenga una longitud mayor a 40 y no presente caracteres especiales, lo cual se define como URL amigable. El resultado del sexto elemento se obtiene al determinar si en todos los documentos obtenidos de los directorios y recolectores está presente la frase “openAccess” en el metadato con etiqueta DC.rights. El último ítem solo implica escoger Sí o No según corresponda.

Posterior al ingreso de la información, se oprime el botón de Guardar, lo cual envía los datos al servidor donde se verifican las URL, que, de ser correctas, devuelve el resultado del criterio; de lo contrario, se muestran los errores. Un ejemplo del resultado del criterio se presenta en la Figura 5.

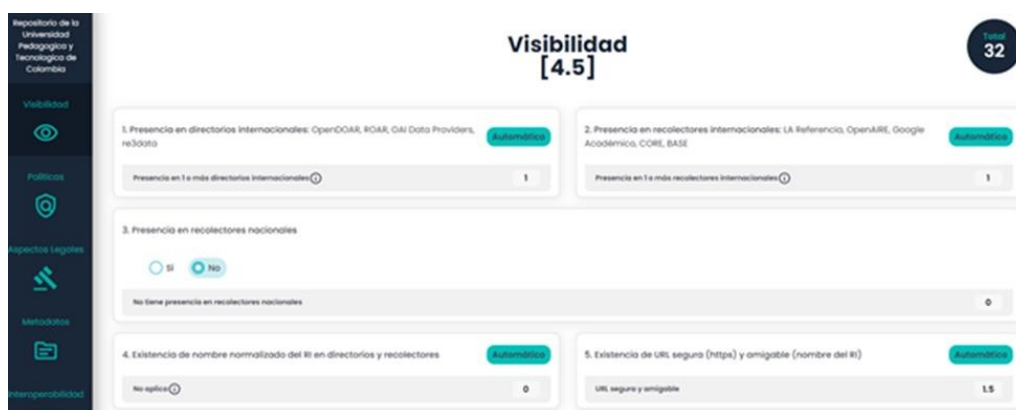


FIGURA 5. Resultado de la evaluación del criterio de visibilidad

Los resultados se muestran por cada ítem, tanto por criterio y como un total, el cual es la sumatoria de todos los criterios evaluados. Además, por cada ítem se muestra un mensaje, la puntuación y, en algunos casos, un botón que abre una ventana que muestra información adicional, la cual corresponde a los documentos o plataformas que no cumplieron con lo evaluado. Algunos ejemplos de la información extra se presentan en la Figura 6.



FIGURA 6. Ejemplo de información adicional a la calificación

A modo de ejemplo, la Figura 7 presenta el resultado del proceso de evaluación del repositorio de la Universidad Pedagógica y Tecnológica de Colombia.



FIGURA 7. Página de resumen de la evaluación

Adicionalmente, en esta página se ubican tres botones que redirigen al detalle del resultado de cada criterio y a la página de inicio, e igualmente existe la opción de descargar el resultado de la evaluación.



## Conclusiones

La evaluación de repositorios es un tema que ha tomado una mayor importancia dentro de la comunidad académica y científica con el auge y crecimiento exponencial de estas plataformas. Como consecuencia de lo anterior, diferentes instituciones, organizaciones e investigadores han realizado propuestas de guías y esquemas para la estandarización de los repositorios. Con el paso del tiempo, se han consolidado guías y documentos que agrupan una mayor cantidad de aspectos. Ejemplos de esto son el Certificado DINI para servicios de documentación y publicación y la Guía para la evaluación de repositorios institucionales de investigación.

Ahora bien, dado el enfoque que se pretendió dar al sistema desarrollado en esta investigación, en la definición de los criterios de la evaluación se tuvo en cuenta la guía de RECOLECTA, pues es la de mayor reconocimiento en la comunidad científica y recibe actualizaciones constantes, como es el caso de la realizada en el 2021, que adicionó nuevos esquemas de metadatos y vocabularios para su contenido.

En la definición de las herramientas y servicios externos se optó por tres diferentes elementos: APIs, web scraping y bases de datos. La primera, en el caso de algunos recolectores y directorios que tienen este servicio; la segunda, principalmente para los puntos en los cuales había que evaluar la información presente en los documentos depositados en los repositorios; y la última, para dos directorios que no tenían API y su información no variaba frecuentemente, por lo cual era más eficiente definir una base de datos con la información de estas plataformas y actualizarla cada cierto tiempo.

Por último, para el desarrollo del aplicativo, se utilizaron los frameworks React y Flask, pues eran del conocimiento previo de los investigadores. Adicional a esto el segundo, al estar basado en Python, posee las librerías más completas para la implementación del web scraping. Aunque no se automatizó la totalidad de los ítems de la evaluación, cabe resaltar que los puntos en los

que esto sí se hizo eran los que mayor tiempo demandaban, pues estaban relacionados con los documentos depositados, y había que revisar cada uno de estos en busca de la información y formatos establecidos en la guía.

En general, el aplicativo permitió disminuir significativamente el tiempo requerido para realizar la evaluación, y esto fue posible gracias a la estandarización del contenido de los documentos de los repositorios. Esto constituye un ejemplo de cómo, por medio de técnicas como el web scraping es posible automatizar procesos o rutinas sobre esquemas y datos previamente establecidos.

## Bibliografía

- AMH, DBV, & ZKI (s.f.) DINI - Deutsche Initiative für Netzwerkinformation. <<https://dini.de>>
- AZORÍN, C., BERNAL, I., GÓMEZ CASTAÑO, J., GUZMÁN PÉREZ, C., LOSADA YÁÑEZ, M., MARÍN DEL CAMPO, R., MARTÍNEZ GALINDO, F. J., MARTÍNEZ POUSA, C., MORILLO MORENO, J. C., & PRATS PRAT, J. (2021). Guía para la evaluación de repositorios institucionales de investigación. Fundación Española para la Ciencia y la Tecnología.
- BOLAÑOS ASENJO, M. Á. (2012). La accesibilidad web en los repositorios institucionales. La UOC a examen. Universitat Oberta de Catalunya.
- BUDAPEST OPEN ACCESS INITIATIVE (2002). Iniciativa de acceso abierto de Budapest. <<https://www.budapestopenaccessinitiative.org/read>>
- DATA CITE METADATA WORKING GROUP (2021). DataCite metadata schema documentation for the publication and citation of research data and other research outputs. <<https://schema.datacite.org/meta/kernel-4.4/>>
- ESCIRE (2019). dPyx - Herramienta de autoevaluación de sistemas de información académica y científica. <<https://dpyx.site>>
- GONZÁLEZ DÍAZ, C., IGLESIAS GARCÍA, M., LLAGUNO, M., & GONZÁLEZ PACANOWSKY, A. (2015). Antecedentes y estado de la cuestión sobre los Repositorios Institucionales de Contenido Educativo (RICE). Departamento de Comunicación y Psicología Social.
- GRUPO DE TRABAJO "PUBLICACIÓN ELECTRÓNICA" (2012). Certificado DINI servicio de documentación y publicaciones. <<https://edoc.hu-berlin.de/handle/18452/18849>>

- JISC (s.f.). OpenDOAR Statistics. Universidad de Nottingham.  
<[https://v2.sherpa.ac.uk/view/repository\\_visualisations/1.html](https://v2.sherpa.ac.uk/view/repository_visualisations/1.html)>
- OPENAIRE (2020). Directrices de OpenAIRE para administradores de repositorios de Literatura v4.  
<<https://guiasopenaire4.readthedocs.io/es/latest/index.html>>
- RED NACIONAL DE REPOSITARIOS DIGITALES DE CIENCIA, TECNOLOGÍA E INNOVACIÓN DE ACCESO ABIERTO (RNRPCTIAA) (2020). Guía Alicia 2.0. RNRPCTIAA.
- SOFTSERVE (2019). Module 7: Accessing DOM with JavaScript. <<https://ppt-online.org/83335>>
- VANDERFEESTEN, M., SUMMANN, F., & MARTIN, S. (2008). Directrices DRIVER 2.0.  
<[https://recolecta.fecyt.es/sites/default/files/contenido/documentos/DRIVER\\_2\\_1\\_Guidelines\\_Spanish.pdf](https://recolecta.fecyt.es/sites/default/files/contenido/documentos/DRIVER_2_1_Guidelines_Spanish.pdf)>

# O desafio do acesso aberto para os povos originários Kaingang na era digital

**CLAUDIA CARMEM BAGGIO**

UNIVERSIDADE FEDERAL DE SANTA CATARINA (UFSC)

**EDGAR BISSET ALVAREZ**

UNIVERSIDADE FEDERAL DE SANTA CATARINA (UFSC)

## **EIXO TEMÁTICO**

Comunicação acadêmica, científica e cultural aberta

## **RESUMO**

Na era da tecnologia digital, compreende-se a importância de fomentar mecanismos para dar visibilidade à produção (técnica, científica e cultural) para a comunidade Kaingang da Aldeia Três Soitas localizada em Santa Maria, RS e para a Escola Estadual Indígena de Ensino Fundamental Augusto Ope da Silva. Como objetivo geral desta pesquisa criou-se uma página na internet para que os docentes e discentes da escola pudessem acessar os materiais pedagógicos de forma simples e organizada. Para a identificação desses materiais pedagógicos, formularam-se os objetivos específicos: investigou-se quais materiais pedagógicos são utilizados pela escola e se estes materiais se encontram em formato digital; quais materiais didáticos se encontram em língua Kaingang ou que abordam a cultura Kaingang e verificou-se se estes materiais estão disponibilizados em acesso aberto. A metodologia adotada foi a pesquisa descritiva e exploratória, desenvolvida a partir da pesquisa bibliográfica e documental, tendo uma abordagem qualitativa, que visa um maior conhecimento sobre a temática dos povos indígenas, elaborada a partir da análise da literatura consultada: livros, artigos de periódicos, teses, dissertações e Portal Capes. Entre os resultados destaca-se a página da escola e também a participação dos gestores da escola, cacique, alunos, professores e moradores da Aldeia na contribuição da pesquisa, onde todos contribuíram para a realização deste estudo. A publicação dos materiais pedagógicos para os Kaingang, na página da escola, contribui para dar visibilidade aos materiais produzidos, possibilitando o compartilhamento de informações e a valorização da língua indígena, contribuindo para a revitalização da cultura, para que a escola indígena cumpra o seu papel de formar cidadãos preparados para fazerem as suas escolhas profissionais e de vida,

qualificando os espaços de suas comunidades, através do acesso aberto à Tecnologia Digital.

#### **PALAVRAS-CHAVE**

Tecnologias Digitais; acesso aberto; Kaingang.

## **1 Introdução**

Na era da informação percebe-se que uma das consequências mais marcantes da sociedade da informação é o crescimento das informações, estas passam a utilizar sistemas mais modernos e para isso se faz necessário ferramentas tecnológicas adequadas para a recuperação e o gerenciamento da informação. Neste contexto, tem-se o conceito de Ciência Aberta “é um conhecimento transparente e acessível, compartilhado e desenvolvido por meio de redes colaborativas” (Vicente-Sáez; Martínez-Fuentes, 2018, p. 7). É um movimento que estimula a transparência do conhecimento científico (Silva; Silveira, 2019), fundamenta-se na premissa de que o conhecimento é um bem público e, portanto, deve estar disponível a todos.

Como objetivo geral desta pesquisa criou-se uma página na internet para que os docentes e discentes da escola pudessem acessar os materiais pedagógicos de forma simples e organizada. A motivação deste trabalho veio da percepção da dificuldade de acesso à materiais didáticos na língua Kaingang, pela comunidade escolar indígena, através da tecnologia digital. Constatada através de uma conversa realizada pela pesquisadora com a diretora da Escola Estadual Indígena de Ensino Fundamental (EEIEF) Augusto Ope da Silva, avaliou-se a necessidade e importância de se ter um web site que facilite a busca de informações, materiais pedagógicos, documentos, legislações, além de divulgar as atividades da escola. Dessa forma criou-se uma página na internet para a Escola da Aldeia Kaingang Três Soitas, localizada na cidade de Santa Maria-RS, mais especificamente para os

docentes e discentes da EEIEF Augusto Ope da Silva, para que estes tenham fácil acesso a estes conhecimentos/informações/documentos, de forma simples e organizada. A delimitação por tal instituição justifica-se por ela ser uma escola diferenciada, multicultural e bilíngue, onde o acesso aos materiais didáticos, específicos da cultura Kaingang, ainda são escassos. Salienta-se que o acesso aberto à informação produzida por este povo originário é um aspecto importante, visto que os conhecimentos produzidos por eles devem ser disponibilizados de forma facilitada, dando visibilidade para quem de direito: o povo Kaingang e a sociedade em geral.

Para a criação desta página, no primeiro momento, foi estabelecido um diálogo com a equipe diretiva da escola para que fosse identificadas as demandas e as expectativas da mesma em relação à criação da página e o principal argumento foi a carência de um espaço organizado com a disponibilidade de materiais para o estudo da língua Kaingang pelos alunos da escola. A partir disso, foram identificados quais materiais pedagógicos são utilizados pela EEIEF Augusto Ope da Silva e que se encontram em formato digital; também foi realizada uma pesquisa em relação aos materiais didáticos em língua Kaingang ou que abordam a cultura Kaingang e que estão disponibilizados em acesso aberto e criou-se a página na Internet da EEIEF Augusto Ope da Silva.

## **2 Desenvolvimento**

Na 41ª Conferência Geral da Unesco, realizada em novembro de 2021, foram discutidas diversas recomendações sobre a Ciência Aberta, movimento que estimula a transparência do conhecimento científico (Silva; Silveira, 2019). Neste documento, se mostra, como uma das recomendações, o diálogo aberto com outros sistemas de conhecimento, como o dos acadêmicos marginalizados, das comunidades locais e dos povos indígenas. Por séculos, criou-se uma ideia distorcida do que é ser indígena, como se essas

comunidades vivessem todas de uma mesma forma, sem contato algum com as áreas urbanas e sem utilizar ferramentas como computadores e celulares. Em consequência disso, cometeu-se o erro de mencionar esses povos quase sempre no passado, sendo que eles vivem também no presente e são parte do nosso futuro.

De acordo com as informações do último Censo Demográfico do Instituto Brasileiro de Geografia e Estatística, a população indígena brasileira é composta por mais de 305 povos. No Brasil estão vivendo aproximadamente 900 mil indígenas, falantes de 274 línguas, desse modo, o povo originário do Brasil, forma, atualmente, uma minoria, representando menos de 0,5% da população brasileira (Instituto Brasileiro de Geografia e Estatística [IBGE], 2013). Uma parte da população indígena vive nas cidades, 324.834 pessoas (IBGE, 2010), estudam em universidades, trabalham com carteira assinada, e estão na internet, produzindo e compartilhando conteúdos muito diversos. Sozinhos, os Kaingang correspondem a quase 50% de toda população dos povos de língua Jê, sendo um dos cinco povos indígenas mais populosos no Brasil.

A mobilização dos povos indígenas, desde a década de 1970, derivou nas políticas e ações que tiveram o seu propósito atingido com a promulgação da Constituição Federal em 1988. Esta reconheceu a identificação da sociedade brasileira como pluriétnica e multicultural. Desta forma, inicia-se um novo modo de se fazer políticas educacionais aos povos indígenas. A Escola indígena passa a ser considerada como comunitária, bilíngue/multilíngue, intercultural, territorializada, específica e diferenciada, intencionando o reconhecimento e a preservação da diversidade cultural e linguística de cada comunidade.

A história da educação Escolar indígena, apesar de não ter uma trajetória tão linear, segundo Ferreira, (2001), pode ser assim compreendida:

A história da educação escolar entre povos indígenas no Brasil pode ser dividida em quatro fases. A primeira situa-se à época do Brasil Colonial, quando a Escolarização dos índios esteve a cargo exclusivo de missionários católicos, notadamente os jesuítas; O segundo momento é marcado pela criação do SPI, em 1910, e se estende à política de ensino da FUNAI e a articulação com o Summer Institute of Linguistics (SIL) e outras missões religiosas; o surgimento de organizações indigenistas não-governamentais e a formação do movimento indígena em fins de 1960 e início de 1970, período de ditadura militar, marcam o início da terceira fase; a quarta fase vem da iniciativa dos próprios povos indígenas, a partir de 1980, que decidem definir e autogerir os processos de educação formal (Ferreira, 2001, p. 72).

O aprendizado por parte dos povos indígenas, das ferramentas de informática e da internet como meios de divulgação de suas culturas e de seus direitos têm recebido maior destaque a cada dia. Thiago (2010) salienta que,

A noção presente no imaginário popular brasileiro de que o índio deixa de ser índio no momento que adota costumes e tecnologias de herança ocidental é contraposta pela prática, cada vez mais difundida nas aldeias indígenas, de utilizar ferramentas de tecnologia de informação exatamente com o intuito de tornar mais eficiente a defesa do estilo de vida e da cultura indígenas.

Dessa forma, o acesso à rede favorece o fator intercultural das tribos. "Criamos uma situação mais justa e igual. Hoje vemos que os índios sofrem problemas de invasão ambiental e ficam muito isolados, não têm meios de se defender", comenta (Thiago, 2010), não será a internet mais uma responsável pela degradação na cultura e hábitos dos índios. "Eles sempre sofreram preconceitos, a rede irá valorizar a cultura entre os povos". Na verdade, a internet já está interligando aldeias indígenas brasileiras.

De acordo com o Censo Escolar 2021, produzido pelo INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira), existem hoje no Brasil 3.359 escolas indígenas. Praticamente metade delas (49%) não possuem saneamento básico, 30% não têm energia elétrica e 75% não têm acesso à internet. Todavia, algumas experiências de políticas públicas têm



conseguido transformar a realidade dessas escolas, agregando de modo proativo criatividade, tecnologias digitais e capacidade de gestão.

A Constituição de 1988 garantiu aos índios o direito de permanecer sendo indígenas. Dessa forma, a legislação atual reconhece a diversidade sociocultural e linguística dos povos indígenas do Brasil.

No que diz respeito à educação escolar, assegurou-se aos povos indígenas o direito a uma Escola específica e diferenciada, o acesso aos conhecimentos de outros povos – ditos universais, pelo uso da língua materna e pela valorização dos conhecimentos e práticas tradicionais dos povos indígenas. São preceitos que se materializam numa escola voltada às atividades do povo de forma específica e diferenciada, no uso de materiais didáticos próprios e construídos com base nos processos próprios de construção de conhecimentos que vem de suas ancestralidades, docência de professores indígenas, membros de suas respectivas comunidades (Ferreira, 2020, p. 85).

Ferreira (2020) destaca que “a educação escolar indígena no Brasil passa a ter diretrizes específicas para sua oferta nas terras indígenas, garantido por lei, que cada povo tenha autonomia para organizar suas escolas de acordo com suas realidades e necessidades”. Sendo assim, isso poderá resultar em experiências bem-sucedidas nos sistemas de educação de cada povo indígena, pois cada um tem uma maneira inerente de educar seus membros. Ferreira (2020) acrescenta que,

Há ainda muito a ser feito no sentido da universalização da oferta de uma educação Escolar de qualidade junto aos povos indígenas no sentido de contribuir para a construção de seus projetos presentes e de futuro, de autonomia e autodeterminação, mas principalmente o reconhecimento no universo das políticas públicas governamentais que visam o atendimento das necessidades básicas de aprendizagem, nos termos propostos pela Constituição Federal de 1988 e as regulamentações que lhe seguem. Porém, mesmo já sendo de conhecimento de todos, ainda é importante lembrar os avanços legais que podemos contabilizar a nosso favor referente à educação escolar.

Após a Constituição de 1988, o Decreto 26, de 04 de fevereiro de 1991, retirou a incumbência exclusiva da FUNAI de administrar processos de educação escolar junto às comunidades indígenas e conferiu ao Ministério da Educação a coordenação e a execução das ações, junto aos Estados e Municípios. A partir da Portaria Ministerial Nº 559, de 16 de abril de 1991, que institui no MEC a Coordenação Geral de Apoio às Escolas Indígenas e o Comitê de Educação Escolar Indígena, se recomenda a criação dos NEIs (Núcleos de Educação Indígena) nas Secretarias Estaduais:

Art.4º - Criar, no Ministério da Educação, uma Coordenação Nacional de Educação Indígena, constituída por técnicos do Ministério e especialistas de órgãos governamentais, organizações não governamentais afetas à educação indígena e universidades, com a finalidade de coordenar, acompanhar e avaliar as ações pedagógicas da Educação Indígena no País.

Art.5º - Estimular a criação de Núcleos de Educação Indígena nas Secretarias Estaduais de Educação, com a finalidade de apoiar e assessorar as Escolas indígenas.

Parágrafo Único - Esses núcleos deverão contar com a participação de representantes das comunidades indígenas locais atuantes na educação, de organizações não governamentais afetas a educação indígena e de universidades (BRASIL, Portaria Nº 559/1991).

Neste contexto, a proposta de uma Escola indígena diferenciada e específica representa um avanço, mas também uma novidade no sistema de educação brasileiro.

Ela exige das instituições de ensino e órgãos públicos responsáveis pela definição de concepções e mecanismos de educação escolar, que os indígenas também façam parte dessas definições, tanto conceituais como de suas práticas para que, de fato, possam incorporar e se incorporar no sistema oficial. Isto representa um desafio ao respeito às particularidades dos povos indígenas, mas que colabora na manutenção de suas especificidades culturais, históricas e linguísticas e propicia novas práticas educativas e de pesquisas em educação indígena, reforçando o que a lei visa assegurar (Ferreira, 2020).

## 2.1 Tecnologias Digitais e os desafios na Era da informação

As tecnologias digitais têm grande impacto sobre o modo como muitos pesquisadores das áreas das ciências humanas realizam e compartilham os resultados de seus estudos. Neste sentido, Rydberg-Cox (2006) salienta que,

Essas ferramentas de busca permitem aos usuários interagirem com os textos de maneiras diferentes que a forma impressa, neste contexto, estes meios digitais abrem novos modos de divulgação e novas formas de apropriação da informação contida nestes textos. Ao mesmo tempo, esses recursos digitais podem mudar radicalmente a audiência e o alcance do trabalho realizado pelos pesquisadores das áreas das ciências humanas.

Para muitas bibliotecas digitais, a crescente disponibilização das tecnologias auxilia a aumentar a produtividade operacional, aprimorando o processo de tomada de decisão; permitindo o desenvolvimento de novos serviços informacionais e atendendo o usuário nas necessidades de busca, acesso e apropriação de informação. Desse modo, as constantes mudanças nas formas de organização tradicional da informação presentes em ambientes informacionais digitais são reflexos da incorporação das Tecnologias de Informação e Comunicação (TICs), como no caso das bibliotecas digitais, que armazenam, preservam, disseminam e permitem o acesso à produção intelectual agregando recursos que possibilitam o processo de construção do conhecimento, a partir da participação colaborativa aplicada em diferentes ambientes.

Nesse sentido, o quadro de competências digitais para os cidadãos, apresenta o vasto conjunto de competências necessárias a todos, desde a *Information Literacy* (expressão traduzida e reconhecida no Brasil como Competência em Informação – CoInfo), passando pela criação de conteúdos digitais, até à segurança e bem-estar nas redes. Desde sua origem, a Competência em Informação tem como objetivo principal desenvolver habilidades para lidar com a informação. O relatório precursor da CoInfo, produzido por Paul Zurkowski em 1974 nos Estados Unidos, indica que:

O documento foi concebido tendo como fundamento as mudanças sociais influenciadas pelo desenvolvimento tecnológico e a constatação de que os cidadãos americanos não possuíam as habilidades para utilizar os recursos tecnológicos disponíveis e que viessem a ser desenvolvidos e apresentava como principal objetivo instruir a população para lidar com a abundância de informações disponíveis que excedem a capacidade humana de avaliar, com a diversidade de procedimentos de busca de informação existentes e com a variedade de rotas de acesso e fontes informacionais mal compreendidas e subutilizadas (Furtado, 2019, p. 65).

Hoje, a ColInfo coopera para a manutenção de uma sociedade democrática, devidamente alicerçada na liberdade humana e na inclusão social (Furtado, 2019), além disso, com a ampliação de desenvolvimento tecnológico e de mecanismos informacionais, os quais repercutem nas atividades cotidianas, nota-se a necessidade de “lidar com informações de forma inteligente e criteriosa [...] um componente essencial de uma educação relevante para a vida no século XXI”, como afirma Haydn (2017) citado por Furtado et al. (2022, p. 25).

Na ocasião da pandemia de Covid-19, viveu-se um cenário catastrófico, em meio à maior crise sanitária e uma grande crise econômica, houve muita desinformação promovida pelo próprio governo federal da época e seus apoiadores. Neste contexto, a tecnologia tornou-se urgente e necessária como um instrumento para verificar a veracidade e autenticidade de informações e documentos, consultar meios de verificação de conteúdo, pesquisar e construir conhecimento. A partir deste exemplo, pode-se ampliar esta ação e aplicá-la a todo e qualquer contexto.

Para Costa, Duqueviz e Pedrosa (2015), “as tecnologias digitais são instrumentos e mediadores da interação humana, e como tais, têm colaborado para mudanças em algumas práticas sociais, tais como a comunicação, a socialização, a organização, a mobilização e a aprendizagem”. Nesse contexto, os povos indígenas têm interesse no uso de tecnologias da sociedade

ocidental contemporânea, principalmente a telefonia móvel e a internet, potencializando o fluxo de informações e a comunicação entre os indígenas, constituindo importantes redes interculturais.

### **2.1.1 Escola Estadual de Ensino Fundamental Augusto Ope da Silva**

Os Kaingang do Rio Grande do Sul possuem 55 Escolas distribuídas em terras já demarcadas e outras ainda em processo demarcatório, atendendo alunos desde a educação infantil, ensino fundamental, até o ensino médio. Ferreira (2020, p. 99) comenta que, dessas escolas, oito oferecem o ensino fundamental e médio e, um Instituto de Educação Indígena forma professor Kaingang em nível de magistério para atuarem nas suas comunidades Kaingang.

Em relação ao nome da Escola, Linda Tuhiwai Smith, intelectual indígena destaca que os nomes indígenas carregam consigo a história do povo, de lugares e eventos. Neste contexto, Ferreira (2020, p. 101), salienta que Augusto foi e ainda é, uma grande liderança Kaingang que sempre esteve na luta por melhores condições para o povo. “Participou do movimento indígena no Brasil, sempre teve como bandeira de luta a retomada e demarcação das terras Kaingang, um grande incentivador e participante na vida dos professores Kaingang na luta por educação escolar indígena”. Defendeu uma educação de qualidade, bilíngue e intercultural. “A cultura viva, a língua viva” (palavras de Augusto Opẽ da Silva). Nesse sentido, mesmo faltando muito o que fazer nas escolas, renomeadas, já representam o pensamento Kaingang.

A Escola Estadual de Ensino Fundamental Augusto Ope da Silva, localizada em Santa Maria, RS, oferece o ensino fundamental e atende alunos da pré-Escola até a EJA (Educação de Jovens e Adultos). Possui cinco turmas multisseriadas: Pré-Escola A e B; Turma 1º aos 3º anos; Turma 4º e 5º anos; Turma 6º aos 9º anos e Turma EJA T3 à T6 - Anos Finais. Nessa Escola trabalham professores indígenas e não indígenas, sendo que a maioria dos professores indígenas atua nos anos iniciais. Nessa Escola, Figura 1, atua uma

professora não indígena na direção e uma professora não indígena na coordenação pedagógica.



FIGURA 1. Escola Estadual Indígena de Ensino Fundamental Augusto Ope da Silva

Fonte: Facebook da EEIEF

A presença de professores Kaingang na escola é um avanço para educação escolar indígena, considerando que num passado próximo os cargos diretivos e à docência eram desenvolvidos somente por professores não indígenas, que alfabetizavam na língua portuguesa. “Isso cumpria os objetivos integracionistas, de negação de valores dos Kaingang, na clara intenção de “civilizar” os indígenas, na busca da integração destes à sociedade nacional e, assim, explorar seu território” (Ferreira, 2020, p. 25).

Em vista disto, Ferreira (2020, p. 33) alerta

para que as tecnologias ajudem as famílias a terem uma vida melhor, ajudem a qualificar suas práticas que vem de geração em geração, tendo a compreensão que a utilização das tecnologias não deve ter a produção como princípio do acúmulo e a riqueza individualizada, mas destina-se a fortalecer relações de complementariedade e reciprocidade entre as pessoas, as famílias, os povos, sem destruir o meio ambiente, a natureza, que lhes disponibiliza parte de sua alimentação numa relação complementar, com a visão da vida tradicional dos

povos indígenas. Uma concepção que parece ser um dos desafios para a educação escolar indígena.

Para Ferreira (2020), a terra é tudo para os indígenas, a partir dessa perspectiva, “a luta por direitos é bem mais ampla e todos os povos procuram estar presentes nas discussões das políticas públicas a seu respeito, entre elas a educação Escolar, que passa a ser umas das prioridades”, pois veem, no artigo 210 da Constituição Federal, o direito de ter uma Escola que respeite suas línguas maternas, bem como os processos próprios de aprendizagem. Mas, além disso, veem a Escola como uma ferramenta potencial de luta. “Nesse sentido, ter a educação escolar como direito é muito mais amplo do que pensar a Escola como um espaço isolado das reivindicações Kaingang” (Ferreira, 2020, p. 39).

Neste contexto, Ferreira (2020, p. 39) comenta que, “com essa compreensão, as lideranças Kaingang participam ativamente na efetivação da educação escolar específica e diferenciada”. Com esse movimento, as Escolas já existentes nas Terras Kaingang vão, aos poucos, ganhando nova forma, outros objetivos, como ações que visam revitalizar a educação e a Escola, baseadas nos ensinamentos ancestrais que sustentarão as lutas e como espaço de construção de pensamento indígena.

Não há indivíduo ou grupo indígena com algum grau de contato que não deseje o acesso, o domínio e a apropriação de uma lista enorme de bens, tecnologias e valores do mundo branco para aperfeiçoar seus conhecimentos e modos de vida, sobretudo se os atuais contextos de pós-contatos são levados em conta. Trata-se, repito, não apenas de tecnologias, mas também de valores e modo de vida (Luciano, 2013, p. 127).

A escola entre os Kaingang é conhecida nos períodos mais antigos da história, como a “escola para índios”, implementada por ações do estado e suas parcerias religiosas, que tinham a intenção de conduzir forçosamente à integração e a assimilação dos Kaingang a chamada “comunhão nacional”, ou

seja, sua extinção como povo culturalmente diferente frente a sociedade nacional branca europeia.

No entanto, a população indígena permanece em situação de vulnerabilidade frente ao avanço violento sobre suas terras, a destruição das florestas e rios, a perda de suas tradições e a redução de sua qualidade de vida. Mesmo com os esforços para assegurar seus direitos fundamentais, essa situação representa um grande desafio também para a educação, em especial para as escolas indígenas.

Nesse sentido, busco a compreensão das razões das políticas que levam os Kaingang a diminuir ou a se envergonhar do uso da língua materna, a perda de costumes, de valores tradicionais da cultura do povo e que propiciaram a desvalorização dos conhecimentos, dos saberes e da pedagogia Kaingang na escola para o índio. Frente a isso, a intensificação da preocupação dos professores, junto com suas lideranças tradicionais e com não indígenas afetos à causa, em refletir e construir uma prática escolar diferenciada e específica, como preconiza a legislação brasileira pós Constituição Federal de 1988 (Ferreira, 2020).

Observa-se que se manifesta de modo mais intenso, a busca de alternativas para a superação das políticas de educação escolar implementadas no passado, colocando em evidência outras práticas e fortalecendo a história do povo.

### **2.1.2 Investimento em formação, infraestrutura e tecnologia na Aldeia Três Soitas e EEIEF Augusto Ope da Silva**

Em meados de 2010, um Programa do Governo Federal, levou energia elétrica para muitas aldeias, a partir disso, as comunidades rurais e tradicionais, avançaram em direção à modernidade, mudaram os hábitos de consumo, com aquisição de eletrodomésticos e melhorias na infraestrutura das comunidades.

A solução encontrada para dar acesso a uma educação mediada pela tecnologia para os povos indígenas Kaingang foi o apoio à escola. Esse espaço



visa oferecer infraestrutura informacional e suporte técnico para os professores. Esta unidade possui internet de banda larga, computadores e impressoras. A Secretaria de Educação do Estado (Seduc-RS), através da 8ª CRE (Coordenadoria Regional de Educação) também oferece formação para os professores para o uso pedagógico desses equipamentos. Além disso, os educadores passam a ter acesso às formações em ambientes virtuais oferecidas pela Secretaria de Educação.

Estreou-se o uso de 10 Chromebooks doados para a Escola Augusto Ope da Silva pelo Governo Estadual no dia 22 de maio de 2022 como pode ser visualizado na Figura 2.



FIGURA 2. Estreia dos Chromebooks na Escola

Fonte: Facebook da EEIEF

Conseguir levar educação para as pessoas que vivem nas Aldeias sempre foi, e continua sendo, um grande desafio para educadores e gestores públicos. Com o objetivo de superar esses desafios, nasce a [página da EEIEF Augusto Ope da Silva](#).

A presença de professores indígenas nas Escolas é imprescindível, pois produz uma melhor compreensão das divergências entre as formas de educar, da educação escolar e da educação familiar ou da educação indígena. “Isso

também possibilita a construção de um Projeto Político Pedagógico das Escolas Indígenas, correspondendo às demandas reais da comunidade, contribuindo para que os conhecimentos oferecidos pela Escola sejam eficazes e sirvam para suas lutas e por seus direitos” (Ferreira, 2020, p. 149).

### **3 Procedimentos metodológicos**

A EEIEF Augusto Ope da Silva tem presente na educação dos alunos, acesso às informações e documentos na página da internet da escola? Para esse estudo foi adotada a pesquisa descritiva e exploratória, desenvolvida a partir da pesquisa bibliográfica e documental, tendo uma abordagem qualitativa, que visa um maior conhecimento sobre a temática dos povos indígenas, elaborada a partir da análise da literatura consultada, composta de material bibliográfico: leis, livros, artigos de periódicos, Portal Capes, teses, dissertações e sites da web relacionados ao tema abordado. Foram selecionados textos sob a temática do acesso aberto, com enfoque no aspecto da transdisciplinaridade e da ciência da informação, visando relacionar ao contexto da Tecnologia de Informação e Comunicação (TICs).

Para a realização desta pesquisa, teve-se a anuência da Direção da Escola, onde a gestora Isabel Cristina Baggio acompanhou e aprovou o desenvolvimento da página da EEIEF Augusto Ope da Silva, localizada na Aldeia Três Soitas, em Santa Maria, RS.

No segundo momento, para a criação da página da EEIEF Augusto Ope da Silva utilizou-se o Google Sites, que é um wiki estruturado e uma ferramenta de criação de páginas da web incluída como parte do pacote gratuito de editores de documentos do Google baseado na web oferecido pelo Google.

No terceiro momento, foram identificados quais materiais pedagógicos são utilizados pela EEIEF Augusto Ope da Silva e que se encontram em formato digital; também foi realizada uma pesquisa em relação aos materiais didáticos

em língua Kaingang ou que abordam a cultura Kaingang e que estão disponibilizados em acesso aberto.

Após, realizou-se uma análise do conteúdo a ser inserido na “Página da Escola”, buscando apresentá-lo de modo simples e organizado, visando inserir as tecnologias digitais na formação dos alunos e professores Kaingang, como ferramenta importante na busca da inclusão social e inclusão digital e na promoção do exercício pleno da cidadania.

#### **4 Considerações finais**

Os Kaingang reconhecem o desafio que se estabelece na atualidade, pautada pela comunicação on-line em massa, para tal desafio, entende-se a escola como colaboradora e lugar estratégico de atuar na educação de crianças e jovens. Nesta perspectiva, diversas iniciativas para o avanço da ciência aberta têm sido realizadas e é necessário que os docentes e discentes das escolas indígenas estejam preparados e colaborem para que essas iniciativas tenham sucesso. Com a criação da página da EEIEF Augusto Ope da Silva oportunizou-se o acesso ao conhecimento para toda comunidade escolar sobre esse importante tema.

Almeja-se que a página da desta escola seja um instrumento que auxilie os professores indígenas a darem visibilidade aos materiais produzidos, possibilitando o compartilhamento de informações e a valorização da língua indígena, contribuindo para a revitalização da cultura, para que a Escola indígena cumpra o seu papel de formar cidadãos preparados para fazerem as suas escolhas profissionais e de vida, qualificando os espaços de suas comunidades, através da Tecnologia Digital.

A importância deste processo reside na possibilidade de deixar registrado, através de fotografias, textos, vídeos, um conhecimento cuja tradição de transmissão é de forma essencialmente oral. Os professores indígenas possuem a visão da organização social e cultural da comunidade, e deste

modo, com a publicação dos materiais produzidos na página da escola, contribuem para a preservação e disseminação da cultura entre os próprios moradores da comunidade e também fora dela.

O desafio presente é provocar os professores a se qualificarem e produzirem material didático e digital em suas línguas e também bilíngue, gerando um impacto positivo no processo de ensino-aprendizagem, na busca de uma educação intercultural, que deve estar associada aos projetos e currículos de manutenção e fortalecimento da identidade dos povos indígenas.

#### 4.1 Sugestões para trabalhos futuros

- Obter parcerias com estagiários da Universidade Federal de Santa Maria/UFSM e a equipe diretiva da Escola para manter atualizada a página da Escola;
- Acompanhar, como voluntária, como está acontecendo a inclusão digital dos indígenas;
- Verificar quais são as contribuições ou os desafios enfrentados pelas lideranças da comunidade indígena e pelos professores, neste processo de inclusão digital.

#### Referências

- BRASIL. (1988). Constituição da República Federativa do Brasil de 1988. Brasília, DF: Presidência da República. <[http://www.planalto.gov.br/ccivil\\_03/constituicao/constituicao.htm](http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm)>
- BRASIL. MINISTÉRIO DA EDUCAÇÃO. (1991). Portaria Interministerial Nº 559, de abril de 1991. Brasília, DF.
- COMISSÃO EUROPEIA. (2022). Combater a desinformação em linha: uma estratégia europeia. <<https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX%3A52018DC0236#>>
- COSTA, S. R. S., DUQUEVIZ, B. C., & PEDROSA, R. L. S. (2015). Tecnologias Digitais como instrumentos mediadores da aprendizagem dos nativos digitais. *Revista Quadrimestral da Associação Brasileira de*

- Psicologia Escolar e Educacional, São Paulo, 19(3), 603-610.*  
<<https://www.scielo.br/j/pee/a/NwwLwRTRTdBDmXWW4Nq7ByS>>
- FERREIRA, B. (2020). ÛN SI AG TÛ PË KI VËNH KAJRÃNRÃN FÃ. O papel da escola nas comunidades kaingang. (Tese de Doutorado). Universidade Federal do Rio Grande do Sul, Faculdade de Educação, Programa de Pós-Graduação em Educação, Porto Alegre.
- FERREIRA, M. K. L. (2001). A educação Escolar indígena: um diagnóstico crítico da situação no Brasil. In: Silva, Aracy Lopes da; Ferreira, Mariana Kawall Leal (org.). *Antropologia, História e Educação: A questão indígena e a Escola*. São Paulo: Global.
- FURTADO, R. L. (2019). A competência em informação no cenário arquivístico: uma contribuição teórico-aplicada. Tese de Doutorado em Ciência da Informação). Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. (2013). Características étnico-raciais da população: classificações e identidades. Rio de Janeiro: IBGE.  
<<https://biblioteca.ibge.gov.br/visualizacao/livros/liv63405.pdf>>
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. (2010). Censo Brasileiro de 2010. Rio de Janeiro: IBGE.
- LUCIANO, G. J. DOS S. (2013). Educação para o manejo do mundo: entre a Escola ideal e a Escola real. Os dilemas da educação Escolar indígena no Alto Rio Negro. Rio de Janeiro: Contra Capa; Laced.
- PORTELA, M. (2013). Humanidades digitais: as humanidades na era da web 2.0. *Rua Larga, 38*. <<https://estudogeral.uc.pt/handle/10316/83568>>
- RYDBERG-COX, J. A (2006). *Digital libraries and the challengers of digital humanities*. Oxford, UK: Chandos Publishing.
- SILVA, F. C. C.; SILVEIRA, L. (2019). O ecossistema da Ciência Aberta. *Transinformação, 31*, e190001.  
<<https://www.scielo.br/j/tinf/a/dJ89vRg94Qxtf6Y7M49Hztr/?lang=pt>>
- SMITH, T. L. (2018). *Descolonizando metodologias: pesquisa e povos indígenas*. Curitiba: Ed. UFPR.
- THIAGO, E. (2010). Brasil: índio, internet e interculturalidade. <<https://pt.globalvoices.org/2010/08/14/brasil-indio-internet-e-interculturalidade/>>
- UNESCO. Open Science.  
<[https://unesdoc.unesco.org/ark:/48223/pf0000379949\\_por](https://unesdoc.unesco.org/ark:/48223/pf0000379949_por)>
- VICENTE-SAERZ, R.; MARTÍNEZ-FUENTES, C. (2018). Open Science now: a systematic literature review for an integrated definition. *Journal of Business Research, 88*, 428-436. DOI: <<https://doi.org/10.1016/j.jbusres.2017.12.043>>

# Rede Sudeste de Repositórios Digitais: a importância do trabalho colaborativo

## CLAUDETE FERNANDES DE QUEIROZ

FUNDAÇÃO OSWALDO CRUZ

INSTITUTO DE COMUNICAÇÃO E INFORMAÇÃO CIENTÍFICA E TECNOLÓGICA EM SAÚDE

[claudete.queiroz@iciict.fiocruz.br](mailto:claudete.queiroz@iciict.fiocruz.br)

### EIXO TEMÁTICO

Comunicación académica y científica en abierto

### RESUMO

Apresenta as ações implementadas pela Rede Sudeste de Repositórios Digitais que foi criada em 2017 e atualmente é coordenada pela Fundação Oswaldo Cruz, com a participação de 84 instituições de Ensino e Pesquisa, que reúne esforços em prol da gestão e da visibilidade da produção científica incluída em seus repositórios. O trabalho tem como objetivo mostrar como as ações colaborativas realizadas pela Rede permitem que os profissionais de informação de cada instituição participante possam relatar e trocar experiências que fortalecerão os Repositórios. Os dados apresentados são sobre os projetos desenvolvidos, reuniões e eventos realizados. A Rede Sudeste faz parte da Rede Brasileira de Repositórios Digitais, compostas pelas regiões Norte, Nordeste, Centro-Oeste, Sul, é coordenada pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT).

### PALAVRAS-CHAVE

Repositórios digitais; Rede Sudeste de Repositórios Digitais; gestão da comunicação científica; rede colaborativa.

## Introdução

Os Repositórios são bases de dados importantes para reunir, disponibilizar e preservar o conteúdo produzido pelas Instituições, permitindo assim, o acesso

à informação e aos objetos digitais. É uma fonte de informação, onde o conhecimento depositado deve ser certificado por profissionais qualificados. Foi a partir do Movimento de Acesso Aberto, que duas estratégias foram definidas para viabilizar o acesso à produção científica: a Via Verde (Green Road) que implantou os repositórios institucionais e o autoarquivamento; e a Via Dourada (Golden Road) que engloba os periódicos de acesso aberto.

Alves e Veiga (2016) relatam que o Acesso Aberto é “um Movimento internacional que visa promover o acesso livre e irrestrito à literatura científica e acadêmica, favorecendo o aumento do impacto do trabalho desenvolvido pelos pesquisadores e instituições, contribuindo, também para a reforma do sistema de comunicação científica”. Dentro desse contexto, podemos entender que a criação, otimização e sustentabilidade dos repositórios, precisam estar alinhados com as políticas de acesso aberto de cada instituição de forma a melhorar o gerenciamento das informações. Os repositórios também permitem reunir em único local a produção dando visibilidade aos trabalhos e o aumento na média de citações. É importante ressaltar que esse sistema precisa possuir requisitos de preservação digital, tais como elaboração de planos, políticas e estratégias, para permitir o acesso aos documentos a longo prazo, tendo em vista a obsolescência tecnológica.

No Brasil, a implantação dos repositórios, ocorreu a partir de duas iniciativas do **Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT)**: a implementação de um projeto piloto com a criação de Repositórios Institucionais e o edital IBICT-FINEP/PCAL/XDBD N° 002/2009. No Projeto piloto foram selecionadas universidades federais para criação dos repositórios, sendo contempladas 27 instituições que receberam kits tecnológicos para operação do software DSpace com treinamento de pessoal e suporte técnico e informacional. Essas iniciativas foram importantes e bem recebidas pelas instituições, acelerando o desenvolvimento dos repositórios institucionais. Diante de um trabalho bem-sucedido realizados pelas instituições que implantaram os RIs, o IBICT decidiu pela criação da Rede Brasileira de

Repositórios Institucionais de Publicações Científicas em Acesso Aberto – RIAA<sup>1</sup> e das Redes Regionais, apoiando as cinco regiões do país: Norte, Nordeste, Sul, Centro Oeste e Sudeste (Assis, 2019).

### **As Redes Regionais de Repositórios Digitais**

O trabalho desenvolvido pelas Redes promove ações individuais e coletivas, que compartilhadas, permitem organizar e estruturar as tomadas de decisões das gestoras responsáveis, permitindo assim, a valorização do conhecimento e o uso das novas tecnologias de informação e comunicação. Essas conexões permitiram criar uma estrutura composta por profissionais de diferentes áreas com valores e objetivos em comum, e que são demonstradas através dos vínculos dos indivíduos e sua percepção pelos resultados alcançados.

A preservação digital da memória institucional é um dos requisitos fundamentais para que o conhecimento produzido pelas Redes, atentos a obsolescência tecnológica, esteja preservado para as gerações futuras, como também possa ser disseminado, compartilhado de forma a estimular o interesse de implantação de repositórios por outras instituições.

Com o objetivo de entender essa integração entre as Redes, descrevemos um pouco do histórico de cada uma para apresentar o que motivou a criação delas. Vale ressaltar que as coordenadoras de cada Rede regional administram todo o trabalho de forma voluntária, sem desconsiderar suas atividades laborais realizadas em suas instituições de origem.

---

<sup>1</sup> Que foi renomeada em 2021 para Rede Brasileira de Repositórios Digitais



## Rede Norte de Repositórios Digitais



Foi a primeira rede a ser criada, a partir da assinatura da [Carta de Belém](#) em 24 de maio de 2014, e atua de forma dinâmica e possui diversas publicações e tutoriais disponíveis em seu [site](#), além de um portal com dados descritivos e canal de vídeos no YouTube. Tem como missão “promover, apoiar e facilitar a adoção da ciência aberta ao conhecimento científico produzido na Região Norte do Brasil, contribuindo para ampliar a visibilidade dos saberes científicos, técnicos e artísticos, disponibilizar e preservar a produção científica e os dados científicos das instituições regionais”. As instituições integrantes são: IFPA, UFRA, MPEG, UFAM, UFT, IFPA, UNIR, IEC, UNIFESSPA e UFPA (Barbalho; Inomata; Galves, 2023).

## Rede Nordeste de Repositórios Digitais



Foi criada em 2015 a partir da iniciativa das Universidade Federal do Rio Grande do Norte (UFRN) e da Universidade Federal de Pernambuco (UFPE). Possui uma [página](#) na internet contendo documentos administrativos com padrões de metadados para teses e dissertações, apresentação de palestras realizadas em encontros da rede, conteúdo com orientação sobre instalação do software DSpace, assim como configurações de repositórios compartilhadas e tem como missão de reunir Instituições de Ensino Superior e de Ensino Profissional e Tecnológico”. As instituições integrantes são: IFS, IFPE, IFBA,

UEPB, UFAL, UFBA, UFERSA, UFPB, UFPE, UFPI, UFRN, UFRPE, UFS, UPE (Almeida, 2022).

### Rede Sul de Repositórios Digitais



Originou-se a partir do compromisso das instituições que promovem a Ciência aberta como um movimento que buscou fomentar, projetar, realizar e, particularmente, democratizar o acesso e uso do conhecimento produzido. Foi criada a partir da assinatura da “Carta de Bagé”, datada de novembro de 2016, assinada em reunião ocorrida na Universidade Federal do Pampa – UNIPAMPA (AMARO, 2019, p. 26). Possui uma [página](#) na internet com os objetivos propostos para atuação em rede.

### Rede Centro-Oeste de Repositórios Digitais



Esta rede foi concebida e conta com a coordenação da Universidade Federal de Goiás, mas a implementação dela ainda não foi efetivada. A expectativa é que se desenvolva como as demais e se torne um incentivo para agregar outras instituições para fortalecer seus Repositórios (Amaro, 2022).

## **A integração e o trabalho colaborativo entre as Redes de Repositórios Digitais**

As redes são compostas por Universidades e Institutos de Pesquisa que atuam como um “conjunto de indivíduos ou organizações interligadas por meio de relações dos mais diversos tipos e passam a ser valorizadas por intensificar a interação, reduzindo o tempo e espaço nas relações” (Verschoore, Balesrin; 2008). Essa associação estratégica tem proporcionado resultados excelentes no desenvolvimento do trabalho colaborativo nas Redes Norte, Nordeste, Sul, Centro-Oeste e Sudeste.

Podemos afirmar que o propósito das Redes está alinhado com as premissas do Movimento do Acesso Aberto e da Ciência Aberta, tendo em vista que “buscam promover a conscientização, o desenvolvimento de estratégias, políticas e infraestruturas que possibilitem o amplo acesso à literatura científica a partir da colaboração mundial e dos resultados das pesquisas científicas” e que precisam ser disponibilizadas para a Sociedade (Almeida, 2022).

Ao elencar os benefícios alcançados pelas Redes, verificamos que as competências e habilidades intrínsecas, viabilizam uma eficiente comunicação entre os seus colaboradores e tornaram-se fundamentais para alcançar melhores resultados nos conteúdos informacionais disponíveis em seus Repositórios. Nesse aspecto, Almeida (2022) pontua que “os Repositórios Digitais (RD) surgiram como uma ferramenta facilitadora no processo de gerenciamento e disseminação da comunicação científica, para reunir, organizar e divulgar estudos e resultados de pesquisas produzidos”.

Dentre os diferentes benefícios da atuação em Redes, destacam-se: compartilhamento de técnicas e procedimentos, visibilidade das instituições e das produções científicas, melhoria dos serviços, gestão da qualidade, fortalecimento das bibliotecas, intercâmbio de conhecimento, implementação de softwares de código aberto e fóruns sobre temas relevantes para a área de

Ciência da Informação em Repositórios Digitais, como Ciência Aberta, Acesso Aberto, Curadoria e Preservação Digital, Direitos Autorais, Métricas, entre outros.

Em 2021, as coordenadoras das Redes e a equipe do IBICT perceberam a necessidade de uma maior integração e decidiram realizar reuniões periódicas para intensificar e fortalecer os trabalhos, bem como propor novas ações para impulsionar o crescimento e a implementação dos Repositórios nas diferentes regiões. Uma dessas ações foi a organização do evento “I Encontro da Rede Brasileira de Repositórios Digitais”, realizado em agosto de 2022, com a transmissão do canal no YouTube da Videosaude da Fiocruz. Durante o evento, diversas palestras abordaram temas importantes, como Ciência Aberta, Dados de Pesquisas, Preservação e Curadoria Digital, Incubadora de Repositórios e Inteligência Artificial. Essa parceria entre as coordenações propiciou a participação de reitores e autoridades das instituições membros no evento, além de contar com palestrantes internacionais e intérpretes de libras, destacando a relevância da acessibilidade. O Encontro teve uma grande participação online, e contou com mais de 2.000 visualizações. Os vídeos estão disponíveis no canal da [Videosaude da Fiocruz](#) e os trabalhos produzidos nos eventos foram depositados no [Arca – Repositório Institucional da Fiocruz](#).

### Rede Sudeste de Repositório Digitais



Os desafios relacionados com a criação, otimização e sustentabilidade dos repositórios e a importância do trabalho cooperativo para o enfrentamento e superação de barreiras, bem como a necessidade de alinhamento das políticas de acesso aberto no Brasil e no âmbito internacional, impulsionou a criação da

Rede Sudeste de Repositórios Digitais<sup>2</sup>, que ocorreu em 3 de outubro de 2017, no Rio de Janeiro, através da assinatura da **“Carta do Rio”**, em que subscreveram 15 instituições de ensino e pesquisa que foram: EMBRAPA, FGV, FIOCRUZ, IBGE, IPEN, INT, SBM, UFF, UFOP, UFJF, UFRJ, UNESP, UNICAMP, UNIFEI e UNIRIO.

### **Metodologia**

A Gestão do Conhecimento é um “processo de organização e distribuição do saber coletivo da organização, de maneira a fazer com que a informação certa chegue à pessoa certa, na hora certa”, sendo fundamental para a otimização dos processos e dos resultados obtidos decorrentes do capital intelectual presente na instituição (Robbins, 2005, p. 242).

A Rede Sudeste, coordenada pela **Fundação Oswaldo Cruz**, é composta atualmente por 84 instituições (Figura 1), incluindo universidades e institutos de pesquisa, tanto públicos quanto privados. A rede e tem promovido diversas ações para a troca de experiências e a colaboração, de forma a enaltecer a importância das Instituições e dos profissionais de informação que trabalham com Repositórios. Esse trabalho, realizado em âmbito regional, fortalece parcerias e facilita o compartilhamento de publicações, garantindo que todo o conteúdo esteja disponível e acessível à sociedade como um todo.

---

<sup>2</sup> De 2017 a 2021 se chamava “Rede Sudeste de Repositórios Institucionais”, e a partir de maio de 2022 foi renomeada para “Rede Brasileira de Repositórios Digitais”

<b>2018</b>	BNDES, BUTANTAN, CNEN, FJP, IBICT, IFSP, INC, INCAPER, INES, UFLA, UFRRJ, UFVJM, UNIFAL/MG, UNIFESP, USP
<b>2019</b>	AERONÁUTICA, ARQUIVO NACIONAL, BIBLIOTECA NACIONAL, CENTRO UNIVERSITÁRIO UNIHORIZONTES, CEPERJ, CPRM, ESCBM, ESCOLA SUPERIOR DE GUERRA, EXÉRCITO BRASILEIRO, FCRB, HOSPITAL SERVIDORES DO ESTADO, IFES, IFRJ, INMETRO, INCA, INSTITUTO DE TECNOLOGIAS DE ALIMENTOS, LNCC, MARINHA DO BRASIL, MAST, NEKI IT, REDARTE/RJ, SANTA CASA DE BELO HORIZONTE, SPU/RJ, UEL, UEMG, UERJ, UFMG, UFSCAR, UFSJ, UFU, UNIVERSIDADE DE VILA VELHA, UVA
<b>2020</b>	BIBLIOTECA MUNICIPAL MARIO DE ANDRADE/SP, COLÉGIO PEDRO II, MUSEU IMPERIAL DE PETRÓPOLIS, SENAI/SP UNIDADE SÃO CAETANO DO SUL, INSTITUTO DE REGISTROS IMOBILIÁRIOS DO BRASIL, MUSEU DA PESSOA/SP, INSPER
<b>2021</b>	SISTEMA FIRJAN, FACULDADE SÃO CAMILO, CETEM, UNIVALE, SBPC ACERVO DIGITAL, SENAC/DN, FESPSP, ARQUIVO PÚBLICO DO ESTADO DO RIO DE JANEIRO (APERJ), UGB/FERP - CENTRO UNIVERSITÁRIO GERALDO DI BIASE
<b>2022</b>	SENAI/SP/ESCOLA SENAI NADIR DIAS DE FIGUEIREDO, ESCOLA DE SAÚDE PÚBLICA DE MINAS GERAIS
<b>2023</b>	PUC/CAMPINAS, FHEMIG/COMPLEXO HOSPITALAR DE URGÊNCIA, UFABC
<b>Total de instituições membros: 84</b>	

FIGURA 1. Instituições membros da Rede Sudeste por ano

As reuniões realizadas pela Rede (Figura 2) desde 2017, tem estabelecido um vínculo entre os profissionais envolvidos, caracterizando assim, os valores institucionais e uma maior percepção da importância da salvaguarda da produção científica dos repositórios envolvidos. Foram 34 reuniões com a presença de 1.282 participantes.

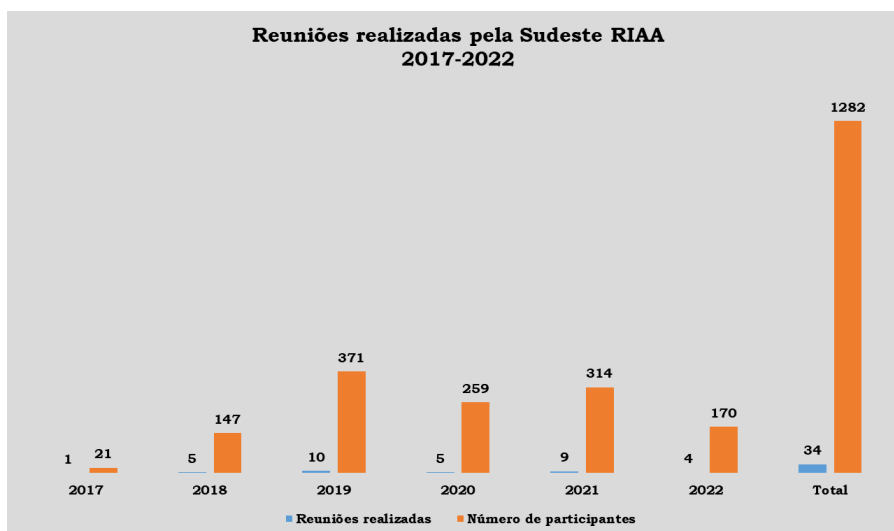


FIGURA 2. Reuniões realizadas 2017-2022

A Rede tem efetuado diversas ações visando a melhoria e a qualidade dos seus repositórios como a criação de um canal de comunicação entre as instituições, através do compartilhamento dos arquivos no [Google Drive](#) e da lista de discussão. A Rede também produziu diversos eventos<sup>3</sup> com temáticas relevantes sobre acesso aberto, ciência aberta, repositórios digitais, curadoria e preservação digital, dados de pesquisa, dentre outros assuntos. Os grupos de trabalho criados permitiram o desenvolvimento de projetos para a capacitação profissional dos membros (Figura 3); o estudo sobre o processo de autoavaliação e certificação, preservação digital e política de indexação para Repositórios.

---

<sup>3</sup> I Encontro da Rede Sudeste de Repositórios Institucionais realizado em 2019 com a presença de 700 participantes; o II Encontro da Rede Sudeste realizado em 2021 e o I Encontro da Rede Brasileira de Repositórios Digitais, realizado em 2022, ambos no formato remoto e que contou com mais de 2.000 visualizações. Os vídeos de 2021 e 2022 estão disponíveis no canal da Videosaude da Fiocruz. Os trabalhos produzidos nos eventos foram depositados no Arca – Repositório Institucional da Fiocruz.

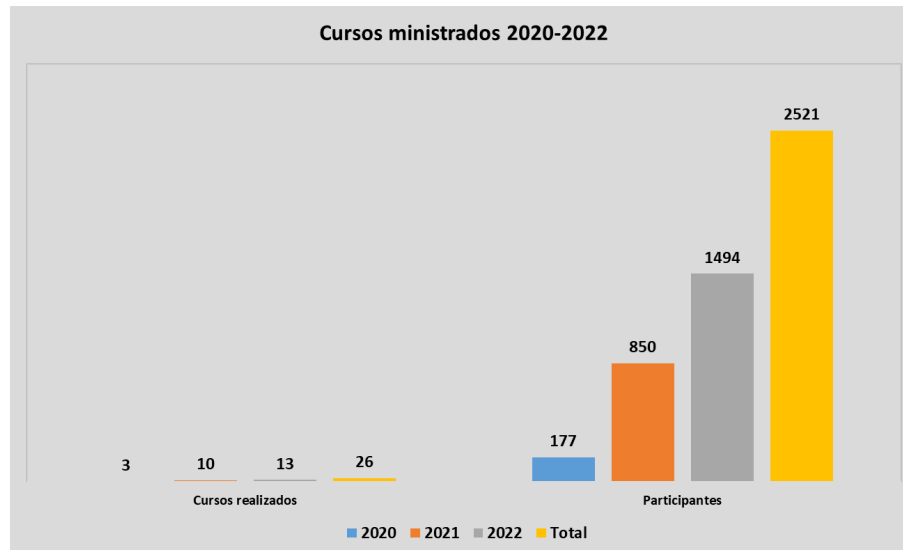


FIGURA 3. Cursos ministrados 2020-2022

Além dessas ações, os membros da Rede têm feito diversas apresentações em eventos e ministrados palestras (Figura 4), referenciando os assuntos que são discutidos durante as reuniões e que são relevantes para os Repositórios. A maioria dos profissionais que compõem a Rede são formados por Bibliotecários que atuam como gestores dos repositórios, e que atendem a Instituição e seus pesquisadores. Também temos outros profissionais como Analistas de Sistemas, Arquivistas e Técnicos em Bibliotecas.

Shintaku e Meirelles (2010) afirmam que a atuação do Bibliotecário como administrador de Repositórios “requer conhecimentos específicos de gestão em vários aspectos como: entender o sistema usado, disseminação, recuperação e organização da informação, alinhamento dos propósitos da instituição, dos usuários e das funcionalidades do repositório”. Portanto, o sucesso da Rede está intrinsecamente ligado a participação desses profissionais que contribuem com seus conhecimentos e experiências vividas.



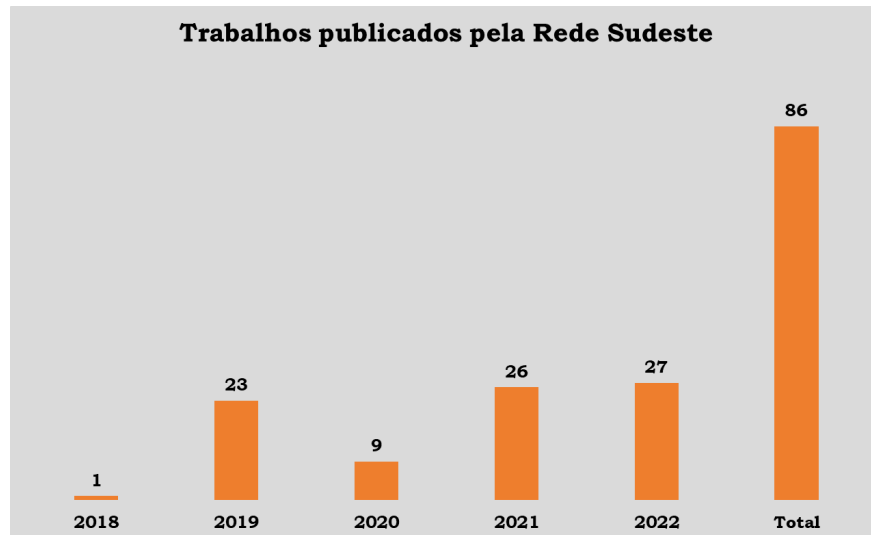


FIGURA 4. Trabalhos publicados 2018-2022

### Resultados parciais e finais em 2022

- Elaboração das atas das reuniões;
- Organização e coordenação dos cursos;
- Atualização do Google Drive da Rede - [link](#)
- Atualização da listagem dos Contatos das Instituições Participantes da Rede;
- Realização de quatro (4) reuniões com 170 participantes e das suas respectivas atas;
- Realização de treze (13) cursos com 1.494 participantes;
- Divulgação na fanpage e Instagram do Arca – Repositório Institucional;
- 65 postagens no Instagram da Rede Sudeste, 44 no Facebook e envio de 44 malas diretas (e-mail);
- Realização do I Encontro da Rede Brasileira de Repositórios Digitais nos dias 09 a 11/08;
- Apresentação de poster no evento da ConfOA – Título: Rede Sudeste de Repositórios Digitais: produção bibliográfica disponibilizada em acesso aberto;

- Depósito das palestras, vídeos e trabalhos apresentados em eventos no Repositório Institucional Arca;
- Permissão para participação de instituições das demais regiões nos cursos organizados;
- Realização parcial do novo diagnóstico da Rede Sudeste;
- Elaboração do relatório de atividades de 2022;
- Elaboração do documento de autoavaliação de Repositórios (projeto Certificação);
- Apresentação de poster no evento da 13ª Conferência Lusófona de Ciência Aberta (ConfOA) – Título: Rede Sudeste de Repositórios Digitais: produção bibliográfica disponibilizada em acesso aberto;
- Ministração de palestra no II Seminário Sobre Periódicos Científicos, Bibliotecas Digitais e Repositório Institucional da Marinha do Brasil – Título: Rede Sudeste de Repositórios Digitais: histórico, atividades e projetos;
- Ministração de palestra no II Encontro Baiano de Bibliotecas Universitárias Públicas – Título: Gestão de dados de Pesquisa, Ciência Aberta e Repositórios;
- Ministração de palestra no evento de lançamento do Repositório Institucional do INCA – Título: A importância do Acesso Aberto e dos Repositórios.

### **Considerações finais**

Conforme proposto para este trabalho, destacamos como estratégicos os principais objetivos da Rede Sudeste que tem contribuído para o sucesso de suas atividades, como: a criação de repositórios digitais, institucionais e temáticos, observando os pressupostos do Acesso Aberto; a busca pelo estabelecimento de uma política de Ciência Aberta para os repositórios de dados de pesquisa; a organização de ações contínuas de capacitação dos

membros; a ajuda mútua na implantação e gerenciamento dos repositórios; a busca por novas instituições que possuam perfil para integrar a rede; o apoio às ações dos grupos de trabalho e o fortalecimento e implantação das Políticas Institucionais para os Repositórios.

A Rede Sudeste de Repositórios Digitais proporciona um ambiente de colaboração e intercâmbio de conhecimento entre os participantes, visando aproximar as instituições para alcançar os critérios de confiabilidade, credibilidade, autenticidade e qualidade das informações depositadas.

## Bibliografía

- ALMEIDA, A. C. L. de (2022). *Conexões em Acesso Aberto: comunidade de prática em Redes Regionais de Repositórios do Brasil*. 166 f. Dissertação (Mestrado em Ciência da Informação)-Natal, UFRN, 2022. <<https://repositorio.ufrn.br/handle/123456789/49978>>
- ALVES, A. DA S., VEIGA, V. S. DE O. (2016) *Repositórios: conceito, tecnologia e aplicação*. Rio de Janeiro, Fiocruz/Icict. 76 p. Trabalho apresentado no Curso de Acesso à Informação Científica e Tecnológica em Saúde. Modalidade: Qualificação. <<http://arca.fiocruz.br/handle/icict/16385>>
- AMARO, B. (2019). *A Via Verde do Brasil e a Rede Brasileira de Repositórios Institucionais de Publicações Científicas em Acesso Aberto – RIAA*. In: Barbalho, C. R. S., Inomata, D. O, Galves, J. M., orgs. 1 ed. *A ciência aberta e seus impactos na Região Norte do Brasil*. Manaus: Edua. p. 11-30.
- AMARO, B. (2022). *A Ciência Aberta e a importância da atuação em rede*. Produção: Rede Brasileira de Repositórios Digitais. [S. l.]. 1 arquivo MP4 (45min03s), son., color.
- ASSIS, T. B. de (2019). *Rede Brasileira de Repositórios e o Impacto dos Trabalhos das Subredes*. In: Encontro da Rede Sudeste de Repositórios Institucionais, 1., 2019, Rio de Janeiro. Anais... Rio de Janeiro: Fiocruz/Icict/UFRJ.
- BARBALHO, C. R. S., INOMATA, D. O., GALVES, J. MACELINO (2019). *A ciência aberta e seus impactos na Região Norte do Brasil*. Manaus: Edua.
- ROBBINS, S. P. (2005). *Comportamento Organizacional*. 11. ed. São Paulo: Pearson Prentice Hall.

- SHINTAKU, M., MEIRELLES, R. F. (2010). Manual do DSpace: administração de repositórios. Salvador: EDUFBA.  
<<https://repositorio.ufba.br/ri/bitstream/ri/769/1/Manual%20do%20Dspace%282%29.pdf>>
- VERSCOORE, J. R., BALESTRIN, A. (2008). Fatores relevantes para o estabelecimento de Redes de cooperação entre empresas do Rio Grande do Sul. *RAC*, 12(4), 1043-1069.  
<<https://www.scielo.br/j/rac/a/Fqfk74PT6Lk7Y8MMhsR8RzG/?format=pdf&lang=pt>>
- WEITZEL, S. R. (2006). O papel dos repositórios institucionais e temáticos na estrutura da produção científica. *Questão*, 12(1), 51-71.  
<<https://seer.ufrgs.br/EmQuestao/article/view/19>>

# Accesibilidad de artículos en formato PDF en revistas científicas de acceso abierto

**ALEXA RAMÍREZ VEGA**

INSTITUTO TECNOLÓGICO DE COSTA RICA

[aramirez@itcr.ac.cr](mailto:aramirez@itcr.ac.cr)

## EJE TEMÁTICO

Comunicación académica, científica y cultural en abierto

## RESUMEN

Actualmente, la discapacidad afecta al 15 % de la población mundial (aproximadamente mil millones de personas), según los datos del primer informe sobre discapacidad de la Organización Mundial de la Salud (OMS). De esos, 645 millones sufren de deficiencias visuales o auditivas, lo que afecta el uso correcto de los documentos académicos disponibles en la web. Los documentos más consultados por estudiantes e investigadores son los artículos científicos disponibles en revistas electrónicas, los cuales en su mayoría se publican en formato PDF (*Portable Document Format*), por lo tanto, es importante garantizar la accesibilidad de esos documentos para personas con discapacidad visual. De esta manera, surge la necesidad de evaluar la accesibilidad de los documentos PDF en las revistas electrónicas de acceso abierto. Para esto se escogieron 150 documentos en PDF de revistas electrónicas disponibles en DOAJ (*Directory of Open Access Journals*) en el área de Comunicaciones y Tecnología de la Información, los cuales fueron evaluados con la herramienta de comprobación de accesibilidad de Adobe Acrobat Pro. De los 150 PDF evaluados, el 100 % presentó algún fallo de accesibilidad. Para los administradores y editores de revistas es importante conocer la situación de accesibilidad de los documentos PDF que se publican en las revistas de acceso abierto con el objetivo de tomar las medidas pertinentes para solventar las fallas y publicar contenidos accesibles para todas las personas.

## PALABRAS CLAVE

Revistas de acceso abierto; PDF accesible; DOAJ; accesibilidad; discapacidad visual.

## Introducción

Actualmente, la discapacidad afecta al 15 % de la población mundial (aproximadamente mil millones de personas), según los datos del primer informe sobre discapacidad de la Organización Mundial de la Salud (OMS) (World Health Organization (WHO), 2011). De esos, 645 millones sufren de deficiencias visuales o auditivas, lo que afecta el uso correcto de los documentos académicos disponibles en la web (World Health Organization, 2015) (World Health Organization, 2014). En general, la discapacidad consiste en una interacción con el entorno, que presenta barreras y limitaciones que afectan a las personas con discapacidad y causan problemas en la interacción. Por lo tanto, la OMS propone hacer posible el acceso a todos los sistemas y servicios convencionales y hacerlos accesibles a personas con cualquier tipo de discapacidad (World Health Organization (WHO), 2011).

Esta situación ha llevado a organizaciones internacionales como el *World Wide Web Consortium* (W3C) a crear recomendaciones y estándares para la Web y el contenido digital, para garantizar a las personas (discapacitadas o no) una participación igualitaria en la Web, como WCAG (Pautas de accesibilidad del contenido web.) [1.0](#) y [2.1](#) (Caldwell, Cooper, Reid, & Vanderheiden, 2008).

Por su parte, algunos autores se han dado a la tarea de evaluar la accesibilidad de PDF en revistas indexadas y los resultados no han sido muy favorables, ya que han encontrado muchas deficiencias en la estructura de los PDF que impide su correcta lectura para personas con discapacidad visual (Nganji, 2015). En (Nganji, 2018) se destaca el esfuerzo de los editores por hacer los PDF accesibles para personas con discapacidad visual, pero hacen énfasis en la inconsistencia y falta de un esquema recomendado para un flujo de trabajo editorial que contemple la evaluación de la accesibilidad de los documentos en PDF que serán publicados. Así mismo, autores como (Angarita Lopez, Fernandez Morales, Nino Vega, Duarte, & Gutierrez Barrios, 2020) han realizado estudios para evaluar la accesibilidad de los sitios web de revistas colombianas del área de humanidades, destacando que ninguna revista

alcanza ninguno de los tres niveles de accesibilidad WCAG 2.1, al no cumplir con todos los criterios esperados en cada nivel.

Como se evidencia en estos estudios, la accesibilidad de los contenidos de las revistas científicas es algo que se ha dejado de lado, y aunque hay revistas que hacen grandes esfuerzos publicando artículos en diversos formatos (HTML, XML) y hasta produciendo audios de los resúmenes y palabras clave, no es suficiente, ya que el formato de mayor publicación es el PDF, por su naturaleza portable y multiplataforma. Por lo tanto, se debe garantizar la accesibilidad (permitiendo a cualquier usuario acceder a la información y los documentos en la Web y utilizarla, independientemente de las condiciones físicas o tecnológicas). Por este motivo, es necesario evaluar la accesibilidad de los documentos PDF en las revistas académicas de acceso abierto y definir recomendaciones que permitan hacerlos accesibles para todos los usuarios.

### **Metodología**

Para la investigación, se tomaron 150 documentos PDF de diferentes revistas electrónicas de acceso abierto disponibles en DOAJ (*Directory of Open Access Journals*) en el área de Tecnologías de la información y comunicaciones.

Para la evaluación, se utilizaron las herramientas de comprobación de accesibilidad de Adobe Acrobat X PRO (comprobación rápida y comprobación completa). Primero, se usó la herramienta de verificación rápida donde se pudo determinar el nivel de accesibilidad del documento. Con esta herramienta se tiene cinco opciones:

- 1) La configuración de seguridad del documento impide el acceso de los lectores de pantalla.
- 2) El documento parece no contener texto. Puede ser una imagen escaneada.
- 3) El documento tiene una estructura lógica pero no es un PDF etiquetado.

- 4) El documento no está estructurado, por lo que el orden de lectura puede no ser correcto.
- 5) No se detectaron problemas de accesibilidad.

Si el documento se ubica en las opciones 3 o 4, se continúa con la verificación completa, de lo contrario, el proceso de evaluación finaliza.

La herramienta de verificación completa crea un informe de accesibilidad, donde se comprueban las siguientes reglas, las cuales se dividen por tipo:

### Documento

- **Permiso de accesibilidad.** Esta primera verificación indica si el documento permite que el texto sea leído y copiado por lectores de pantalla para efectos de convertirlo en audio. Es decir, verifica si el acceso a herramientas de accesibilidad está permitido.
- **PDF de imágenes.** Indica cuando el contenido del PDF no es texto legible. Esto podría suponer que sea un documento escaneado y por lo tanto el texto no sea legible.
- **PDF etiquetado.** Se refiere a los documentos que no están etiquetados para especificar el orden de lectura lógica.
- **Orden de lectura lógica.** Esta regla se verifica manualmente. Asegúrese que el orden de lectura que se muestra en el panel Etiquetas coincida con el orden de lectura lógico del documento.
- **Idioma principal.** La configuración del idioma del documento en un PDF permite que algunos lectores de pantalla cambien al idioma apropiado. Esta comprobación determina si se especifica el idioma del texto principal para el PDF.
- **Título.** Informa si hay un título en la barra de título de la aplicación Acrobat.



- **Contraste.** Comprueba si el documento contiene elementos que no sean accesible para las personas con baja visión o daltónicos.

## Contenido

- **Contenido etiquetado.** Verifica si todo el contenido del documento está etiquetado. Esta comprobación se puede hacer manualmente, asegúrese de que todo el contenido del documento esté incluido en el árbol de Etiquetas.
- **Anotaciones etiquetadas.** Esta regla verifica si todas las anotaciones están etiquetadas. Esta comprobación se puede hacer manualmente, asegúrese de que las anotaciones como comentarios y marcas editoriales (como insertar y resaltar) estén incluidas en el árbol de Etiquetas del PDF.
- **Orden de tabulación.** Debido a que las tabulaciones se usan a menudo para navegar por un PDF, es necesario que el orden de las tabulaciones sea paralelo a la estructura del documento.
- **Codificación de caracteres.** Especificar la codificación ayuda a que los visores de PDF presenten texto legible a los usuarios.
- **Elementos multimedia etiquetados.** Esta regla comprueba si todos los objetos multimedia están etiquetados.
- **Parpadeo de la pantalla.** Los elementos que hacen que la pantalla parpadee (por ejemplo, las animaciones y los scripts), pueden causar convulsiones en personas que padecen epilepsia fotosensible. Estos elementos también puede ser difíciles de ver cuando se amplía la pantalla.
- **Secuencias de comandos (scripts).** El contenido no puede depender de secuencia de comandos (scripts) a menos que las tecnologías de soporte puedan tener acceso al contenido y su funcionalidad.

- **Hipervínculos accesibles.** Para que los lectores de pantalla puedan tener acceso a las direcciones URL, en el PDF deben tener hipervínculos activos correctamente etiquetados.

### Formularios

- **Campos de formulario etiquetados.** Todos los campos del formulario deben estar etiquetados y formar parte de la estructura del documento.
- **Campos con descripción.** todos los campos de formulario necesitan una descripción de texto.

### Texto alternativo

- **Texto alternativo de figuras.** las imágenes del documento deben tener texto alternativo y que este sea significativo a la figura que corresponde.
- **Texto alternativo anidado.** Los lectores de pantalla no leen el texto alternativo en elementos anidados.
- **Texto alternativo de otros elementos.** Este informe comprueba contenido distinto de figuras que necesita texto alternativo (como multimedia, anotación o modelo 3D).

### Tablas

- **Filas en tablas.** Esta regla comprueba si cada TR en una tabla es un elemento secundario de Tabla, THead, TBody o TFoot.
- **TH y TD en tablas.** Esta regla verifica que en la estructura de tabla, TH y TD sean elementos secundarios de TR.
- **Encabezados.** Esta opción verifica que todas las tablas del PDF tengan un encabezado.

- **Regularidad.** Para ser accesibles, se verifica que las tablas contengan el mismo número de columnas en cada fila y de filas en cada columna.
- **Resumen en tablas.** Se verifica que la tabla contenga resumen, aunque es opcional, puede mejorar la accesibilidad.

## Listas

- **Elementos de lista.** La comprobación informa de si cada Elemento de lista (LI) es un elemento secundario de Lista (L). Cuando esta comprobación de la regla tiene errores, la estructura de esta lista no es correcta.
- **Lbl y Lbody.** Comprueba si las listas tienen la siguiente estructura: un elemento de lista debe contener Elementos de lista. Y Elementos de lista solo puede contener Elementos de etiqueta y Elementos de cuerpo de lista. Cuando esta comprobación de la regla tiene errores, la estructura de esta lista no es correcta.

## Encabezados

- **Anidamiento adecuado de encabezados.** Esta regla comprueba los encabezados anidados. Cuando esta comprobación tiene errores, los encabezados no están correctamente anidados.

Cada una de estas reglas se encuentra explicada con detalle en (Adobe Acrobat, 2020), donde se expone el objetivo de la comprobación, así como las maneras de solventar el problema si es que ha fallado en alguna de ellas.

## Resultados

De los 150 documentos, el 100% de ellos presentaron problemas de accesibilidad. El 2% de los documentos tenían una configuración de seguridad que no permitía la evaluación con la herramienta de verificación. El 98% de los documentos restantes eran de texto y si fue posible escanear su contenido, sin embargo, éstos presentaron algún problema de accesibilidad. En la figura 1 se muestran la cantidad de PDF con fallos en cada uno de los tipos de reglas de accesibilidad evaluados. Ahí se evidencia que la mayor falla se dan en elementos del documento, texto alternativo y tablas.

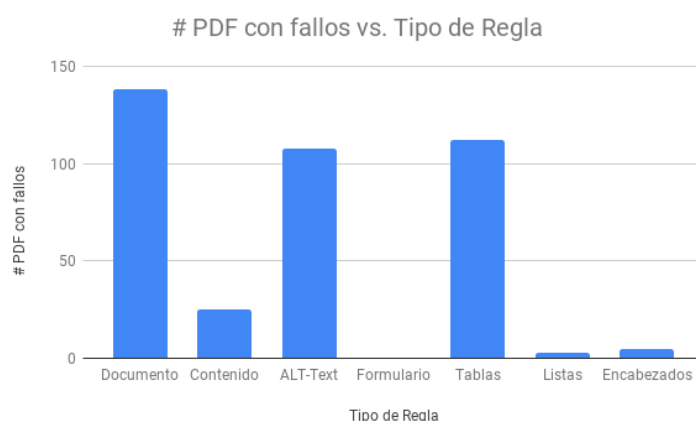


FIGURA 1. Cantidad de PDF con fallos según cada tipo de regla

Cabe mencionar que el tipo la comprobación para elementos de formulario es cero, debido a que, por el tipo de documento evaluado, ninguno incluye estructuras de formulario. Por su parte, en el cuadro 1 se detalla la cantidad de PDF que tuvieron fallas en cada regla evaluada con la herramienta de Accesibilidad de Adobe Pro.

<b>TIPO DE REGLA</b>	<b>REGLA</b>	<b># PDF CON FALLOS</b>
Documento	Permiso de Accesibilidad	2
	PDF de imágenes	3
	PDF etiquetado	6
	Orden de lectura lógica	11
	Idioma principal (LANG)	49
	Título	52
	Contraste	15
	TOTAL	138
Contenido	Contenido etiquetado	5
	Anotaciones etiquetadas	0
	Orden de tabulación	15
	Codificación de caracteres	0
	Elementos multimedia etiquetados	0
	Parpadeo de la pantalla	0
	Secuencias de comandos (scripts)	0
	Hipervínculos accesibles	2
	TOTAL	22
Formulario	Campos de formulario etiquetados	0
	Campos con descripción	0
	TOTAL	0
Texto alternativo	Texto alternativo de figuras (ALT-TEXT)	98

<b>TIPO DE REGLA</b>	<b>REGLA</b>	<b># PDF CON FALLOS</b>
	Texto alternativo anidado	0
	Texto alternativo de otros elementos	10
	TOTAL	108
Tablas	Filas en tablas	9
	TH y TD en tablas	12
	Encabezados	63
	Regularidad	28
	Resumen en tablas	-
	TOTAL	112
Listas	Elementos de lista	2
	Lbl y Lbody	1
	TOTAL	3
Encabezado	Anidamiento adecuado de encabezados	5
	TOTAL	5

CUADRO 1. Cantidad de PDF con fallas según cada regla evaluada

Los problemas más críticos indicados en el cuadro 1 se relacionaron con etiquetas faltantes, un 34 % no tenía la etiqueta de idioma principal (LANG), un 74 % omitía el texto alternativo en imágenes u otros elementos (ALT-TEXT) y el 26 % que los incluían no era texto coherente o significativo, por su parte el 36 % omitió la etiqueta de título (TITLE). Estas fallas causan problemas con los lectores de pantalla y hacen que los documentos sean ilegibles para los usuarios con cualquier tipo de discapacidad visual.

## Conclusiones y recomendaciones

La apertura de las revistas electrónicas significa que cualquier persona puede acceder al contenido y éste debe considerar sus particularidades de cualquier país, cualquier edad, utilizando cualquier tecnología (móvil o de escritorio) o funciones de accesibilidad como lectores de pantalla, zoom, colores invertidos. Por esta razón, los editores de revistas deben garantizar la accesibilidad de todos los documentos publicados en sus sitios. Además, los desarrolladores podrían incluir una verificación de accesibilidad automática en las plataformas de revistas para ayudar a los editores en esta labor.

Adicionalmente, los índices y bases de datos que evalúan las revistas científicas se centran en aspectos formales de la publicación (periodicidad, políticas, etc) y de contenido (revisión por pares, contenido original), pero este tipo de requerimientos de accesibilidad para personas con algún grado de discapacidad visual ha quedado por fuera. Esto permite identificar la importancia de exigir a las revistas que sus documentos y sus sitios web sean accesibles para todos. Y mientras más accesible sea para una persona con discapacidad, más fácil y amigable será para cualquier persona sin ninguna discapacidad.

De esta manera, dado que muchos de los fallos de accesibilidad encontrados pueden ser fácilmente solventados y para garantizar la accesibilidad de los PDF de revistas de acceso abierto se recomienda:

- Diagramar la revista siguiendo lineamientos de diseño sobre accesibilidad, como:
  - Fondo blanco y letras oscuras. Evitar el uso de colores pastel o de bajo contraste que dificulten la lectura.
  - Usar tipografía denominada “palo seco” (carecen de remates o extensiones en los trazos).
  - Tamaño óptimo de letra para lectura, entre 12 y 18 puntos.

- Estructura de encabezados lógica correcta. Es decir, uso adecuado de H1, H2, H3.
- Evitar la publicación de documentos PDF cuyo texto sea imágenes. De igual forma, verificar que los datos incluidos en tablas sean texto y no imágenes.
- Verificar que las figuras incluidas dentro de los documentos incluyan la etiqueta ALT-TEXT y que este texto indique información útil para el lector. Es decir, que describa la imagen lo mejor posible.
- Verificar que el PDF tenga la etiqueta LANG correctamente asignada.
- Verificar que el PDF sea legible por lectores de pantalla.
- Hacer comprobación de accesibilidad utilizando Adobe Pro u otra herramienta. Si su documento es elaborado desde Word también puede comprobar la accesibilidad usando las herramientas disponibles en ese software [8] (Moreno, Martínez, & González, 2014).
- Seguir guías de evaluación y adecuación de materiales digitales, como los expuestos por los autores (López González, Palmeros y Ávila, & Coeto Calcáneo, 2020) y de documentos PDF accesibles como los citados en (Álvarez Lacambra & Moreno López, 2021).

## Bibliografía

- ADOBE ACROBAT. (2020). Create and verify PDF accessibility, Acrobat Pro. Retrieved December 2022. <https://helpx.adobe.com/acrobat/using/create-verify-pdf-accessibility.html?trackingid=KACNN#DocTitle>
- ÁLVAREZ LACAMBRA, A., & MORENO LÓPEZ, L. (2021). Publicar documentos accesibles en PDF: disponibilidad no es lo mismo que accesibilidad. Sedic. Sociedad Española de Documentación e Información Científica.
- ANGARITA LOPEZ, R., FERNANDEZ MORALES, F., NINO VEGA, J., DUARTE, J., & GUTIERREZ BARRIOS, G. (2020). Accesibilidad de las revistas



colombianas del área de humanidades bajo las pautas WCAG 2.1. *Espacios*, 41(4), 18.

CALDWELL, B., COOPER, M., REID, L., & VANDERHEIDEN, G. (2008). Web content accessibility guidelines (WCAG) 2.0 W3C Recommendation. World Wide Web Consortium (W3C).

LÓPEZ GONZÁLEZ, V., PALMEROS Y ÁVILA, G., & COETO CALCÁNEO, I. (2020). Guía para la evaluación y adecuación de materiales digitales accesibles. *Emerging Trends in Education*, 3(5).

MORENO, L., MARTÍNEZ, P., & GONZÁLEZ, Y. (2014). Guía para elaborar documentación digital accesible. Recomendaciones para Word, Power Point y Excel de Microsoft Office 2010. CENTAC.

NGANJI, J. (2015). The Portable Document Format (PDF) accessibility practice of four journal publishers. *Library & Information Science Research*, 37(3), 254-262.

NGANJI, J. (2018). An assessment of the accessibility of PDF versions of selected journal articles published in a WCAG 2.0 era (2014-2018). *Learned Publishing*, 31(4), 391-401.

WORLD HEALTH ORGANIZATION (WHO). (2011). World Report on Disability. WHO.

WORLD HEALTH ORGANIZATION. (2014, August). Visual impairment and blindness - Fact Sheet No 282.  
<<http://www.who.int/mediacentre/factsheets/fs282/en/>>

WORLD HEALTH ORGANIZATION. (2015, Marzo). Deafness and hearing loss - Fact sheet No 300.  
<<http://www.who.int/mediacentre/factsheets/fs300/en/>>

# Disponibilização de dados de pesquisa em repositórios universitários da América Latina

## **JAQUELINE INSAURRIAGA SILVEIRA**

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

[jaqueline.silveira@ufrgs.br](mailto:jaqueline.silveira@ufrgs.br)

## **SAMILE ANDRÉA DE SOUZA VANZ**

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

[samilevanz@terra.com.br](mailto:samilevanz@terra.com.br)

## **CATERINA MARTA GROPOSO PAVÃO**

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

[caterina@cpd.ufrgs.br](mailto:caterina@cpd.ufrgs.br)

### **EIXO TEMÁTICO**

Datos abiertos

### **RESUMO**

O trabalho analisa a disponibilidade de dados de pesquisa em repositórios universitários da América Latina. Metodologicamente, o estudo abrange uma pesquisa documental onde identifica no re3data repositórios registrados como sendo para dados de pesquisa, para posteriormente selecionar os que são institucionais de universidades. Descreve os repositórios quanto ao armazenamento e disponibilidade de dados de pesquisa, restrições de acesso e licenças de uso. Constatou-se que o re3data não representa uma fonte de pesquisa fidedigna para recuperação de repositórios de dados, uma vez que as informações registradas são autodeclarações das instituições. Dessa forma, a pesquisa sofreu interferência devido a escolha do Diretório como ferramenta de coleta de dados. Considera-se que o objetivo do presente estudo não foi totalmente alcançado, uma vez que dos 20 repositórios analisados, somente seis são exclusivos para dados de pesquisa.

#### **PALAVRAS-CHAVE**

Dados de pesquisa; disponibilização de dados; repositório institucional; América Latina.

#### **ABSTRACT**

The paper analyzes the availability of research data in university repositories in Latin America. Methodologically, the study covers documentary research where it identifies re3data repositories registered as being for research data, to later select those that are institutional of universities. It describes the repositories in terms of storage and availability of research data, access restrictions and use licenses. It appears that re3data does not represent a reliable research source for retrieving data repositories, since the information recorded are self-declarations by the institutions. Thus, the research suffered interference due to the choice of the Directory as a data collection tool. It is considered that the objective of the present study was not fully achieved, since of the 20 repositories analyzed, only six are exclusive for research data.

#### **KEYWORDS**

Research data; data availability; institutional repository; Latin America.

## **1 Introdução**

A Ciência Aberta é entendida como um movimento que envolve práticas e iniciativas de abertura da ciência em diferentes âmbitos. Dentre estas, destaca-se para o presente estudo, a abertura dos dados de pesquisa com suas implicações para a ciência, oportunizando sua reprodutibilidade e desenvolvimento de novos estudos; para os pesquisadores, potencializando a troca de conhecimento, otimizando tempo e recursos de pesquisa; e para a sociedade, por meio de descobertas que impulsionam políticas públicas.

Ao possibilitar o compartilhamento e reuso, a abertura de dados de pesquisa proporciona a reprodutibilidade da ciência, além de fornecer agilidade aos processos científicos e otimização de recursos financeiros para novos estudos. A partir desse entendimento, atores envolvidos nos contextos científico, financeiro e político da ciência (pesquisadores, comunidade científica,

instituições, governos e agências de fomento), empenham esforços para que, pesquisas financiadas com recursos públicos estejam acessíveis ao público, em formato digital sem restrição ou com restrição mínima (Torino, Roa-Martinez & Vidotti, 2020).

No entanto, o compartilhamento de dados de pesquisa requer iniciativas que garantam a sua disponibilização e integridade, de forma permanente ao longo do seu ciclo de vida. Ao encontro dessa exigência, tem-se, os repositórios digitais e especialmente os Repositórios de Dados de Pesquisa (RDP) como sistemas de informação adequados para o depósito, acesso e uso dessa tipologia documental, uma vez que dispõem das necessárias condições técnicas e tecnológicas de armazenamento e preservação.

A região Latino Americana, segundo Amaro (2021), é tida como modelo de produção científica em acesso aberto, tendo em vista a forte atuação das editoras universitárias. Especificamente no Brasil, o lançamento da SciELO em 1998 e a tradução do software livre Open Journal Systems pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) em 2003 foram os principais motivos pelo qual o acesso aberto se difundiu rapidamente (Vanz & Silva Filho, 2019). Diante desse contexto de reconhecimento da grande adesão da América Latina às publicações em acesso aberto, pode-se inferir que, em relação a abertura e disponibilização de dados de pesquisa, a Região repete o mesmo protagonismo.

Partindo dessa inferência, este estudo objetiva conhecer as iniciativas de abertura de dados de pesquisa dos países da América Latina, por meio da disponibilização em repositórios universitários de dados de pesquisa, a existência ou não de restrições de acesso, e a atribuição de licenças de uso. As próximas seções apresentam os procedimentos metodológicos, uma apresentação descritiva dos repositórios encontrados seguida pela discussão destes resultados e conclusão.

## 2 Procedimentos metodológicos

Para a identificação dos repositórios no âmbito da América Latina, realizou-se uma busca no [Registry of Research Data Repositories \(re3data\)](#) no período de 12 a 22 de fevereiro de 2023, utilizando-se do filtro “País”. Foram recuperados 59 repositórios sediados em seis países, com o seguinte quantitativo de repositórios: Argentina (oito); Brasil (18); Chile (dois); Colômbia (doze); México (dezessete); Panamá (dois). Dentre estes, realizou-se uma triagem selecionando somente os repositórios institucionais de universidades, resultando neste quantitativo: Argentina (quatro); Brasil (quatro); Chile (um); Colômbia (nove) e México (dois). O repositório argentino Humadoc Repositorio<sup>1</sup> não foi analisado por não atender aos critérios metodológicos estabelecidos para o presente estudo, ou seja, analisar repositórios institucionais de universidades, não fazendo parte do estudo aqueles vinculados a órgãos específicos das instituições.

Cada um dos repositórios registrados no re3data foi visitado para coleta de dados, caracterizando a pesquisa documental. A obtenção das informações sobre acesso, restrições e licenças de uso se deu por meio de análise dos conjuntos de dados (datasets) e seus metadados recuperados por meio de estratégias de busca e mecanismos de visualização disponibilizadas pelos repositórios. Quando não se obteve sucesso na recuperação de datasets, buscou-se informações nos sites e políticas de gestão desses repositórios. Importante ressaltar que o re3data não coleta informações sobre os repositórios, sendo estas registradas por seus gestores. Dessa forma é possível a existência de outras iniciativas que por não estarem registradas no Diretório, não fazem parte deste estudo.

---

<sup>1</sup> Repositório da Facultad de Humanidades de la Universidad Nacional de Mar del Plata.

Nos casos em que a coleta de dados e informações diretamente no repositório não atendeu a demanda da pesquisa, foi realizado contato com o setor responsável pelo repositório por e-mail ou Whatsapp.

### **3 Apresentação e análise dos resultados**

A seguir, apresenta-se os repositórios identificados no re3data, por país e suas características principais.

#### **3.1 Argentina**

Identificaram-se quatro repositórios universitários, apresentados a seguir.

##### **3.1.1 Producción Académica UCC**

O Producción Académica UCC é o repositório institucional da Universidad Católica de Córdoba (UCC), fundada em 1956. Gerenciado pelo Sistema de Bibliotecas da Universidade, reúne a produção acadêmica e científica gerada por professores e pesquisadores da Universidade, objetivando sua preservação a longo prazo e divulgação, com a garantia de acesso livre e aberto (UCC, 2021). Disponibiliza em seu website as políticas de: conteúdo; de acesso; de depósito; de metadados; de preservação; de serviços; e de confidencialidade e privacidade.

Embora sua Política de conteúdo contemple os dados de pesquisa nos tipos documentais, ao realizar-se uma busca em sua plataforma, recuperou-se somente um dataset, o qual é disponibilizado em acesso aberto, sem restrições e nem exigência de identificação de usuário, correspondendo a um conjunto de documentos XML que dão respaldo a um caderno de trabalho (workingpaper - WP).

Não foi possível localizar diretrizes específicas para dados de pesquisa, sendo assim, infere-se que estes obedeçam às orientações destinadas aos outros tipos de documentos. O acesso ao Repositório é aberto e todos os

conteúdos depositados estão sob a proteção de licenças Creative Commons (CC). Os autores devem conceder ao Repositório uma licença de distribuição não exclusiva sob a qual será feito o arquivamento, a divulgação e a preservação dos documentos. Para os conteúdos, os autores são encorajados a atribuir uma Licença Atribuição - Não Comercial - Sem Trabalhos Derivados (CC BY-NC-ND 4.0).

Visando confirmar a recuperação de somente um dataset nas buscas realizadas no Repositório, uma vez que sua interface de pesquisa apresentou instabilidade recorrente em diferentes momentos da pesquisa, buscou-se contato via e-mail com o Sistema de Bibliotecas da UCC. Obteve-se, por meio da Diretora do Sistema, a confirmação do armazenamento de somente um dataset e, da não existência de uma política específica para dados de pesquisa, justificado pelo fato da Universidade estar dando os primeiros passos ao encontro da gestão de dados de pesquisa na Instituição.

### **3.1.2 Repositorio de Datos Académicos UNR (RDA-UNR)**

O RDA-UNR, lançado no ano de 2022, além de ser o repositório de dados de pesquisa da Universidad Nacional de Rosario (UNR), é o primeiro repositório aberto de dados de pesquisa de uma universidade argentina. Também é um repositório de autopublicação, onde os usuários carregam, gerenciam e publicam seus dados, contando com a ajuda de curadores e guias de utilização. Com o objetivo de dar visibilidade, promover a reutilização e garantir a preservação dos dados, o Repositório além de receber dados de pesquisa gerenciados por professores e alunos pesquisadores da Universidade, permite também que pesquisadores de outras instituições que publicam artigos em periódicos do Portal de Revistas UNR depositem os dados correspondentes às suas investigações (Universidad Nacional de Rosario, 2022).

Disponível no website do Repositório, sua política de gestão abrange as diretrizes de: Acordo de uso geral do RDA-UNR; Contrato de liberação de dados de pesquisa; Contrato de uso de dados; Licenças de uso de dados;

Compromisso de citação de dados; Dados aceitos no repositório; Quantidade e tamanho dos dados; Dados sensíveis e confidenciais; e de Preservação.

A gestão e uso dos dados de pesquisa no Repositório incorrem na aceitação das condições do Contrato de liberação de dados de pesquisa, pelos autores de dados e, do Contrato de uso de dados, pelos usuários que acessam, baixam e utilizam os dados de pesquisa armazenados no RDA-UNR. Aos dados de pesquisa é concedida uma Licença Atribuição 4.0 Internacional (CC BY BY4.0), permitindo compartilhamento e adaptação dos dados, com sua devida citação de autoria. O RDA-UNR organiza-se em 17 Dataverses vinculados a faculdades e órgãos da UNR.

### **3.1.3 Repositorio Institucional de la Universidad Católica de Salta (UCASAL)**

O repositório institucional da Universidad Católica de Salta (UCASAL) tem como objetivo a promoção do acesso aberto à produção científico-tecnológica gerada pela Instituição. O acesso ao Repositório se dá pela página do Sistema de Bibliotecas da Universidade, e armazena em formato digital a documentação acadêmica e científica gerada e publicada pela instituição, disponibilizando seu conteúdo em acesso aberto.

O Repositório está organizado em quatro grandes coleções: Colección Tesis y Trabajos Finales; Colección Investigación; Colección Histórica Institucional e Colección Archivo Fotográfico. De acordo com a Resolução Reitoral 1017/17, os dados de pesquisa devem ser armazenados na Colección Investigación, porém estes, não estão contemplados na descrição de conteúdo dessa coleção. Em contato com o Sistema de Bibliotecas da Universidade obteve-se a confirmação de que no presente momento não há conjunto de dados processados no Repositório.

Para o licenciamento de conteúdos, a Resolução Reitoral 1017/17 promove o uso das Creative Commons como forma de manter a autoria da produção científica da Universidade, facilitando seu uso para fins acadêmicos e de investigação (Universidad Católica de Salta [UCASAL], 2017).



### **3.1.4 Servicio de Difusión de la Creación Intelectual (SEDICI)**

O Servicio de Difusión de la Creación Intelectual (SEDICI) é o repositório institucional da Universidad Nacional de La Plata (UNLP), tendo por objetivo indexar, preservar e disponibilizar em acesso aberto toda a produção intelectual produzida na Universidade, dentre estes, datasets. O Repositório divide-se em quatro grandes coleções: Publicações acadêmicas e científicas; Coleções patrimoniais; Dados de pesquisa; e Documentos institucionais.

A coleção Dados de pesquisa contém 85 datasets, e verificou-se que os dados estão linkados com as publicações relacionadas. A disponibilização dos dados se dá em acesso aberto sem exigência de identificação e registro de usuário. Em relação às licenças de uso, salvo indicação contrária, são atribuídas Creative Commons, permitindo compartilhar, copiar e redistribuir o material em qualquer meio ou formato para qualquer finalidade, mesmo comercialmente, desde que seja dado o crédito apropriado e este não seja modificado (CC BY-ND).

## **3.2 Brasil**

A pesquisa no re3data recuperou 18 repositórios, sendo quatro destes universitários, a saber:

### **3.2.1 Base de Dados Científicos da Universidade Federal do Paraná – BDC/UFPR**

A Base de Dados Científicos da Universidade Federal do Paraná (BDC/UFPR) tem por objetivo reunir os dados científicos utilizados em pesquisas publicadas pela comunidade da Universidade Federal do Paraná (UFPR) em teses, dissertações, artigos de periódicos e outros materiais bibliográficos. Lançada em setembro de 2017, a BDC/UFPR é gerenciada pelo Sistema de Bibliotecas da Universidade em parceria com o Centro de Computação Científica e Software Livre (C3SL) da Instituição, e representa a primeira base de dados de pesquisa de uma universidade pública brasileira (Estevão, 2018).

A coleção de dados conta com 65 documentos publicados entre os anos de 2017 e 2022. Realizou-se o acesso aleatório à coleção de dados de pesquisa, constatando-se sua disponibilização em acesso aberto com a atribuição de licenças Creative Commons CC0 1.0 Universal (CC0 1.0), ou, uma Atribuição-NãoComercial-SemDerivados 3.0 Brasil (CC BY-NC-ND 3.0 BR).

### **3.2.2 Open Research Data@PUC-Rio**

O Open Research Data@PUC-Rio é um agregador para facilitar o acesso à dados de pesquisa, entre outros conteúdos digitais no Repositório Maxwell. O Maxwell é o repositório institucional da Pontifícia Universidade Católica do Rio de Janeiro (PUC-RIO), desenvolvido e mantido pelo Laboratório de Automação de Museus, Bibliotecas Digitais e Arquivos (LAMBDA), laboratório da Vice-reitoria Acadêmica (VRAc) da Universidade (Pontifícia Universidade Católica do Rio de Janeiro [PUC-Rio], 2022). Com o objetivo de compartilhar dados que foram desenvolvidos e/ou modificados no âmbito da Universidade para que estes sejam utilizados em estudos na Instituição, o Open Research Data@PUC disponibilizou em 2015 seu primeiro conjunto de dados de pesquisa. Para submissão de conteúdo é necessário que, pelo menos, um dos autores pertença à comunidade PUC-Rio e que os dados sejam disponibilizados em acesso aberto. Todos os datasetes submetidos devem ser licenciados com uma licença Creative Commons, que segundo informações disponibilizadas no site, é uma atribuição (CC BY-NC-SA 4.0).

### **3.2.3 Repositório Institucional UNESP**

O Repositório Institucional UNESP pertence à Universidade Estadual Paulista, que sob a responsabilidade do Grupo Gestor da Política do Repositório Institucional UNESP (GRI-UNESP) tem por objetivo armazenar, preservar, divulgar e disponibilizar em acesso aberto à produção intelectual dos pesquisadores e alunos da Universidade (Universidade Estadual Paulista [UNESP], 2022). Dentre os documentos armazenados disponibiliza também dados de pesquisa e data paper. Em pesquisa realizada utilizando-se o filtro

'Tipo de documento', recuperou-se 83 datasets, publicados entre os anos de 2008 e 2023. Dentre estes, há dados em acesso aberto e dados com restrição de acesso. Em relação à atribuição de licenças de uso, não foi possível localizar nenhuma informação no site do Repositório e nem nos metadados dos datasets.

#### **3.2.4 Repositório de Dados de Pesquisa da Unicamp (REDU)**

O Repositório de Dados de Pesquisa da Unicamp (REDU) é o repositório institucional de dados abertos de pesquisa da Universidade Estadual de Campinas (Unicamp). Criado em 2020 e coordenado por um comitê científico e técnico composto por bibliotecários, profissionais de TI e pesquisadores representantes dos usuários, o REDU armazena, preserva, e disponibiliza os dados de pesquisa produzidos na Universidade, permitindo seu compartilhamento e reutilização. De acordo com a política de Ciência Aberta da Unicamp, o Repositório apresenta cinco dataverses (Artes e Humanas; Exatas; Ciências Aplicadas; Biomédicas; Tecnológica), onde encontram-se depositados 138 datasets e 933 arquivos publicados entre os anos de 2018 e 2023. A disponibilização dos dados de pesquisa se dá por prerrogativa em acesso aberto com uma atribuição de Domínio Público (CC0), contudo pode ser necessário o embargo de dados, mas com disponibilização dos seus metadados em acesso aberto.

### **3.3 Chile**

No re3data recuperaram-se dois repositórios chilenos, contudo somente um deles é universitário, o **Repositorio de Datos de Investigación de la Universidad de Chile**. Este, é um repositório institucional que objetiva a preservação, divulgação e disponibilização de dados de pesquisa gerados por sua comunidade de pesquisadores e acadêmicos. É composto por 19 dataverses, 50 datasets e 334 arquivos de dados publicados entre os anos de 2019 e 2023. A disponibilização dos arquivos se dá em acesso aberto com indicação de

atribuição de Domínio Público (CC0) nos Termos de uso dos dados, e também em acesso restrito, exigindo a identificação de usuário.

### **3.4 Colômbia**

A pesquisa no re3data retornou 12 repositórios, sendo somente nove pertencentes à universidades, os quais são apresentados na sequência.

#### **3.4.1 Repositório Institucional Intellectum (RII)**

O Intellectum é o repositório institucional da Universidad de La Sabana, tendo sua criação aprovada pela Comissão de Assuntos Gerais do Conselho Superior na Lei no 1370, de 6 de dezembro de 2012. Criado para gerenciar, preservar e divulgar a produção intelectual, científica, cultural e histórica de sua comunidade universitária, tem sua administração a cargo da Direção da Biblioteca Octavio Arizmendi Posada. O RII é composto por seis comunidades, sendo elas: Trabajos de Grado; Recursos Educativos Digitales Abiertos (REDA); Memoria Institucional; Publicaciones Académicas; Archivo Histórico Cipriano Rodríguez Santa María; Artículos de Investigación Indexados; e, Otros Documentos.

A disponibilização dos documentos se dá em acesso aberto e acesso restrito e para licenciamento de uso são atribuídas licenças Creative Commons adaptadas para o país. Por padrão, a indicação é de atribuição (CC BY-NC-SA 4.0), salvo quando orientação diferente. Não foi possível verificar a disponibilização de dados de pesquisas no RII, uma vez que não se obteve sucesso na recuperação destes no Repositório. Da mesma forma, em suas políticas não existem diretrizes para essa tipologia documental. Ainda buscando confirmar a existência ou não de dados de pesquisa no RII, entrou-se em contato via WhatsApp com a Biblioteca gestora do Repositório. Obteve-se a informação de que não há dados de pesquisa armazenados no RII, visto estarem ainda em processo de criação da coleção Datos de Investigación.

### **3.4.2 Repositorio Universidad Autónoma de Bucaramanga (UNAB)**

O Repositório UNAB é o repositório institucional da Universidad Autónoma de Bucaramanga (UNAB), para fins de armazenamento, preservação e divulgação da produção intelectual, científica, cultural e histórica da sua comunidade universitária. É composto por seis comunidades de documentos, sendo elas: Investigación; Normatividad institucional; Publicaciones académicas; Recursos multimedia; Tesis y disertaciones; e, Trabajos de grado. Não se obteve resultado ao se efetuar uma busca por dados de pesquisa na plataforma, utilizando-se o filtro 'Tipologia documentária'. Da mesma forma, não foram localizadas as políticas de gestão do Repositório.

### **3.4.3 Repositorio Institucional Universidad Distrital Francisco José de Caldas (RIUD)**

O RIUD é o repositório institucional da Universidad Distrital Francisco José de Caldas destinado a armazenar sua produção intelectual. Conta com cinco comunidades de documentos: Documentos RIUD; Producción Editorial Universidad Distrital Francisco José de Caldas; Recursos de Enseñanza; Tesis y Disertaciones Académicas; Trabajos de grado. Não foi possível localizar dados de pesquisa por meio de busca aleatória em suas comunidades documentais.

Para as licenças de uso de conteúdo, sua política explicita que os titulares de direitos autorais cedem à Universidade uma licença não exclusiva, limitada e gratuita. A licença é aplicada por 10 anos prorrogáveis indefinidamente, de acordo com os direitos de autor, o qual pode rescindi-la mediante aviso prévio antes de findar o prazo inicial ou de renovação (Universidad Distrital Francisco José de Caldas, 2022). A disponibilização dos documentos no RIUD se dá em acesso aberto e acesso restrito e, para licenciamento de uso desses, segundo sua política, atribui-se licenças Creative Commons, não ficando claro contudo, se a atribuição das licenças fica a cargo do depositante ou da Instituição.

#### **3.4.4 Repositorio Institucional de la Universidad Santo Tomás**

O Repositorio Institucional de la Universidad Santo Tomás é destinado ao armazenamento, preservação e divulgação de toda produção acadêmica e administrativa da Universidade. Gerencia 10 comunidades de documentos digitais: Banco de objetos virtuales de aprendizaje; Centro de Recursos para el Aprendizaje, la Investigación y la Docencia; Colegios de la Orden de Predicadores; Documentos CRAI-USTA; Gestión Institucional; Grupos de Investigación; Libros; Publicaciones seriadas; Responsabilidad Social Universitaria; Tesis, Trabajos de Grado y Disertaciones Académicas. Realizou-se uma busca aleatória em suas comunidades para identificação de datasets, contudo não se obteve êxito, e dessa forma não se pode verificar como ocorre a disponibilização destes. O que se pode observar em relação aos documentos acessados aleatoriamente dentro das Comunidades é que estes são disponibilizados em acesso aberto com atribuição das licenças Creative Commons.

#### **3.4.5 Repositorio UNAL**

O Repositorio UNAL é o repositorio institucional da Universidad Nacional de Colombia (UNAL), gerenciado por sua **Biblioteca Digital**, com a finalidade de conservar e difundir a produção acadêmica e científica da Universidade. O Repositório disponibiliza acesso aberto aos documentos armazenados e disposto nas seguintes comunidades: Tesis y Disertaciones; Libros y capítulos de libro; Documentos e tutoriales para publicar; Proyectos temáticos; Trabajos de grado; General; Revistas UNAL. Dentro da tipologia documentária armazenada nas comunidades, não foi possível identificar datasets. Em relação ao licenciamento dos conteúdos depositados, os autores concedem ao Repositório uma licença não exclusiva, limitada e gratuita. A disponibilização dos documentos se dá em acesso aberto com atribuição de licenças Creative Commons.

### **3.4.6 Repositorio Institucional Universidad de Antioquia**

O Repositorio Institucional Universidad de Antioquia é coordenado pelo Sistema de Bibliotecas da Universidade. Sua política de gestão expressa a importância do acesso aberto aos conteúdos digitais como um meio de democratização do conhecimento. Armazena artigos de revistas científicas e acadêmicas; livros; capítulos de livros; trabalhos de graduação, mestrado e doutorado; trabalhos de eventos; workingpapers; relatórios de pesquisa; material audiovisual; e documentos históricos e patrimoniais. Estes conteúdos são aceitos somente após passarem por uma revisão por pares.

O Repositório conta com nove comunidades, sendo elas: Biblioteca Virtual de Antioquia; Editoriales; Información institucional; Investigaciones; Patrimonio bibliográfico y documental; Producción docente; Producción estudiantil; Trabajos de posgrados; e Trabajos de pregrado. Analisou-se a descrição de conteúdo das Comunidades e em nenhuma delas consta datasets, contudo, a Universidade possui uma Política Institucional de Acesso Aberto, onde discorre que o Plano de Desenvolvimento Institucional (2017-2026) se compromete com a política e estratégias da Ciência Aberta e que a Universidade procurará incorporar gradualmente em seu repositório, dados de pesquisas, recursos educativos, software e outros.

### **3.4.7 Repositorio Institucional Universidad Autónoma de Manizales (REPOUAM)**

O REPOUAM, repositório institucional da Universidad Autónoma de Manizales (UAM), tem o compromisso com o acesso aberto a todos os tipos de conteúdo digital, incluindo dados de pesquisa. Estruturado em comunidades e coleções, o REPOUAM contempla as comunidades: Institucional (composta por documentos institucionais); Investigación; Recursos educativos; Sello editorial (composta por produções editoriais da Universidade; Tesis y disertaciones; e Trabajos de grado. Em uma busca aleatória no Repositório, foi possível localizar na Comunidade Investigación, um dataset relacionado a um estudo

sobre contratura muscular. A disponibilização deste dataset se dá em acesso aberto licenciado por uma atribuição (CC BY-ND 4.0).

#### **3.4.8 Repositorio Digital UTB**

O Repositorio Digital UTB é o repositório institucional da Universidad Tecnológica de Bolívar (UTB). Disponibiliza todos os documentos produzidos no âmbito da Universidade por meio das comunidades: Fototecas; Investigaciones; Producción académica; e Sello editorial. Não se obteve sucesso na recuperação de dados de pesquisa em sua plataforma e na tentativa de localizar suas políticas para esclarecer sobre as diretrizes e orientações sobre a tipologia de conteúdo aceita para depósito. Em relação à disponibilização de acesso aos documentos em geral armazenados, pode-se constatar que essa se dá em acesso aberto e acesso restrito e, os documentos são licenciados com atribuições Creative Commons.

#### **3.4.9 Repositorio de datos de investigación de la Universidad del Rosario**

Lançado em 2019, o Repositório tem por objetivo preservar e disponibilizar acesso aos dados de pesquisa resultante de projetos da Universidade, visando seu compartilhamento e reuso. Suas normativas seguem a Política Institucional de Gestão de Dados de Pesquisa, orientada pela Portaria Reitoral nº 1.576, de março de 2019. A Portaria orienta sobre a tipologia de dados reconhecidos para depósito no Repositório, permitindo a inclusão de novos tipos conforme necessidade futura. Da mesma forma, discorre sobre dados anônimos, anonimizados, pessoais, públicos, semiprivados, privados, sensíveis, e dados pessoais de crianças e adolescentes. No momento da coleta de dados para este estudo, o Repositório contava com 118 registros de dados de pesquisa dispostos em diversas coleções. Muitos datasets estão licenciados com Creative Commons Domínio Público (CC0) e Atribuição-NãoComercial 4.0 Internacional (CC BY-NC 4.0), outros têm acesso restrito para consulta, sendo disponibilizados somente seus metadados.



### **3.5 México**

Recuperaram-se no re3data 17 repositórios mexicanos, contudo somente dois universitários.

#### **3.5.1 Portal de Datos Abiertos da Universidad Nacional Autónoma de México (UNAM)**

O Portal de Datos Abiertos da UNAM funciona como um ponto de acesso online para busca de acervos universitários autorizados para uso, reutilização e redistribuição gratuita desde que respeitados os Termos de Uso Gratuito de Dados Abertos da UNAM. Contém dados, objetos digitais e camadas geoespaciais de coleções biológicas, trabalhos artísticos, música, medicina veterinária, projetos universitários, entre outros (Universidad Nacional Autónoma de México [UNAM], 2022). O Portal tem uma estrutura onde, inicialmente, apresenta os números estatísticos das coleções, contando com 2.106.655 registros nas diversas coleções. A disponibilização de acesso aos datasets se dá em acesso aberto, mas alguns conjuntos de dados contam com algumas restrições que demandam encaminhamento de solicitação de acesso ao coordenador do projeto. Todos os conteúdos são licenciados com Creative Commons Atribuição 4.0 Internacional (CC-BY 4.0), exceto quando se indique o contrário.

#### **3.5.2 Repositorio de Acceso Abierto (RIAA)**

O RIAA é o repositório institucional de acesso aberto da Benemérita Universidad Autónoma de Puebla (BUAP), que permite consultas às informações científicas, tecnológicas e de inovação produzidas na Universidade, assegurando a disponibilidade e acessibilidade a esse conteúdo, sem restrições, desde que respeitada a observância das atribuições das licenças Creative Commons (Benemérita Universidad Autónoma de Puebla [BUAP], 2018b).

De acordo com a descrição de conteúdo do RIAA-BUAP, este abriga principalmente artigos, livros, teses, documentos de eventos científicos e

outros materiais voltados para a produção de conhecimento; edições especiais, livros incunábulo, materiais antigos, manuscritos e coleções fotográficas (BUAP, 2018a). Na descrição de conteúdo das comunidades e coleções não se identificou a indicação de dados de pesquisa. Contudo, a comunidade EcoBUAP (Ecosistema de Aprendizaje Abierto BUAP) dá acesso ao [Repositorio EcoBUAP](#), o qual disponibiliza a comunidade Conjunto de Datos. No entanto ao realizar-se busca por assunto nesta comunidade, não se obteve resultados.







#### 4 Considerações sobre os repositórios

Surgidos no contexto do Open Access, no início dos anos 2000, os repositórios institucionais (RI) são uma extensão do conceito de repositórios digitais desenvolvidos na década de 1980, originalmente para o armazenamento de softwares de computadores (Macgregor, 2023). Suas características específicas estão diretamente relacionadas ao conteúdo documental a ser armazenado. De forma geral Bicknese (2003-2004), na década de 2000, classificou os repositórios em dois modelos, sendo o primeiro modelo destinado à coleta de documentos de uma única disciplina, e o outro modelo destinado à coleta de todo material produzido por uma faculdade ou universidade, referindo-se a este como sendo os repositórios institucionais. Posteriormente, à medida que os repositórios digitais foram evoluindo tecnicamente e assim diversificando seus propósitos e funcionalidades, desencadeou-se o surgimento de novas e distintas tipologias (Macgregor, 2023). Na tipologia proposta por Macgregor (2023) têm-se alguns exemplos em funcionamento, são os repositórios: de preprints (ex., EmeRI, 2020, cooperação entre a ABEC e o IBICT, com apoio da Unesco); de objetos de aprendizagem (MIT OpenCourseWare, 2001); de dados governamentais (Portal Brasileiro de Dados Abertos); agregadores (LA Referencia, 2012, Acordo de Cooperação entre países da América Latina); mega-repositórios (repositórios de escopo generalista, a exemplo do Zenodo, 2013); e os de dados de pesquisa. Macgregor (2023) ainda inclui em sua proposta os “repositórios confiáveis” que

seriam aqueles que satisfazem a requisitos rigorosos de confiança, lhes sendo atribuídos um selo como o “CoreTrustSeal”.

Para os repositórios digitais de dados de pesquisa, segundo Sayão e Sales (2016) há um certo consenso na literatura quanto a sua tipologia, caracterizando-os como: repositórios institucionais de dados de pesquisa, vinculados a instituições acadêmicas ou institutos de pesquisas, com a finalidade de armazenar toda produção científica gerada por seus pesquisadores; repositórios disciplinares de dados de pesquisa, destinados ao arquivamento de material de áreas específicas do conhecimento; repositórios multidisciplinares de dados de pesquisa, reúnem datasets de diversas áreas do conhecimento; e os repositórios de dados de pesquisa orientados por projetos, os quais recebem coleções de dados oriundas de projetos de pesquisa ou resultantes de resolução de problemas específicos.

A partir das tipologias apresentadas e das análises realizadas, têm-se que dos 59 repositórios recuperados no re3data, somente 20 são repositórios digitais institucionais vinculados a universidades, e dentre esses 20, somente seis são repositórios institucionais de dados de pesquisa, sendo os outros 14, repositórios institucionais acadêmicos, como mostra o Quadro 1.



Países	Total de repositórios no re3data	Repositórios Universitários	Repositório de Dados de Pesquisa
 Argentina	8	4	1
 Brasil	18	4	3
 Chile	2	1	1
 Colômbia	12	9	1
 México	17	2	0
 Panamá	2	0	0
<b>Total</b>	<b>59</b>	<b>20</b>	<b>6</b>

QUADRO 1. Quantitativo de repositórios universitários e de dados recuperados no re3dataFonte

Fonte: Dados da pesquisa







Pode-se observar no Quadro 1, que o Brasil apresenta o maior número de repositórios institucionais registrados no Diretório, 18 repositórios, seguido pelo México com 17, Colômbia com 12, Argentina com oito, e Chile com somente dois repositórios. Quando se seleciona somente os repositórios institucionais de universidades, a Colômbia aparece com o maior número de repositórios registrados no re3data, seguida pelo Brasil e Argentina, depois México e, por último o Chile. Ao passo que dentre os repositórios universitários, identificando-se somente os que são de dados de pesquisa, o Brasil aparece novamente, com o maior quantitativo registrado no Diretório, seguido pela Argentina, Chile e Colômbia com somente um repositório cada país. Por meio de divulgação de grupos de pesquisas e instituições de pesquisa, sabe-se que no Brasil além dos três repositórios registrados no re3data, estão em fase de planejamento distintas iniciativas de implementação de repositórios de dados de pesquisa (PAVÃO et al., 2022).

No Quadro 2, a seguir, encontra-se uma síntese sobre os 20 repositórios recuperados no re3data, podendo-se identificar: o país de origem; o ano de sua criação; sua tipologia; software utilizado para o gerenciamento das coleções; resultado da busca por dados de pesquisa na plataforma; e como se dá a disponibilização de acesso a esses dados.

REPOSITÓRIO	PAÍS	ANO <sup>2</sup> CRIAÇÃO	TIPO DE REPOSITÓRIO	SOFTWARE	RECUPERAÇÃO DADOS DE PESQUISA	TIPO DE ACESSO AOS DADOS	LICENÇA ATRIBUÍDA
<b>Producción Académica UCC</b>		2013	Institucional acadêmico	EPrints	sim	Acesso aberto	(CC BY-NC- ND 4.0)
<b>Repositorio de Datos Académicos UNR</b>		2022	Dados de pesquisa	Dataverse	sim	Acesso aberto	(CC BY4.0)

<sup>2</sup> Informações coletadas nos sites dos repositórios e em pesquisas na *web*.

REPOSITORIO	PAÍS	ANO <sup>2</sup> CRIAÇÃO	TIPO DE REPOSITÓRIO	SOFTWARE	RECUPERAÇÃO		LICENÇA ATRIBUÍDA
					DADOS DE PESQUISA	TIPO DE ACESSO AOS DADOS	
<b>(RDA-UNR)</b>							
<b>Repositorio Institucional UCASAL</b>		2017	Institucional acadêmico	PhpMyBibli (PMB)	não	Não identificado	Não especificado
<b>SEDICI (Servicio de Difusión de la Creación Intelectual)</b>		2003	Institucional acadêmico	DSpace	sim	Acesso aberto	(CC BY-ND)
<b>Base de Datos Científicos da Universidade Federal do Paraná</b>		2017	Dados de pesquisa	DSpace	sim	Acesso aberto	(CC0 1.0), ou (CC BY-NC-ND 3.0 BR)
<b>Open Research Data@PUC-Rio</b>		2015	Dados de pesquisa	Maxwell	sim	Acesso aberto	(CC BY-NC-SA 4.0)
<b>Repositorio Institucional UNESP</b>		2016	Institucional acadêmico	DSpace	sim	Acesso aberto e fechado	Não especificado
<b>Repositorio de Dados de Pesquisa da Unicamp (REDU)</b>		2020	Dados de pesquisa	Dataverse	sim	Acesso aberto e fechado	(CC0)
<b>Repositorio de Datos de Investigación de la Universidad de Chile</b>		2019	Dados de pesquisa	Dataverse	sim	Acesso aberto	(CC0)
<b>Repositorio Institucional Intellectum (RII)</b>		2012	Institucional acadêmico	DSpace	não	Não identificado	(CC BY-NC-SA 4.0)
<b>Repositorio Universidad Autónoma de Bucaramanga</b>		2021	Institucional acadêmico	DSpace	não	Não identificado	Não especificado
<b>Repositorio Institucional Universidad Distrital - RIUD</b>		2016?	Institucional acadêmico	DSpace	não	Não identificado	Não especificado
<b>Repositorio Institucional de la</b>		?	Institucional	DSpace	não	Não	Não

REPOSITORIO	PAÍS	ANO <sup>2</sup> CRIAÇÃO	TIPO DE REPOSITÓRIO	SOFTWARE	RECUPERAÇÃO		LICENÇA ATRIBUÍDA	
					DADOS DE PESQUISA	TIPO DE ACESSO AOS DADOS		
<b>Universidad Santo Tomás</b>			acadêmico				identificado	especificado
<b>Repositorio UNAL</b>		?	Institucional acadêmico	DSpace	não	Não identificado	Não especificado	
<b>Repositorio Institucional Universidad de Antioquia</b>		2008	Institucional acadêmico	DSpace	não	Não identificado	Não especificado	
<b>Repositorio Institucional Universidad Autónoma de Manizales (REPOUAM)</b>		?	Institucional acadêmico	DSpace	sim	Acesso aberto	(CC BY-NC-SA 4.0)	
<b>Repositorio Digital UTB</b>		2017?	Institucional acadêmico	DSpace	não	Não identificado	Não especificado	
<b>Repositorio de datos de investigación de la Universidad del Rosario</b>		2019?	Dados de pesquisa	Dataverse	sim	Acesso aberto e fechado	(CC0) e (CC BY-NC 4.0),	
<b>Portal de Datos Abiertos da Universidad Nacional Autónoma de México (UNAM)</b>		2016	Institucional acadêmico	Não identificado	sim	Acesso aberto e fechado	(CC-BY 4.0)	
<b>Repositorio de Acceso Abierto (RIAA)</b>		2022?	Institucional acadêmico	DSpace	não	Acesso aberto	Não especificado	

QUADRO 2. Repositórios, país, ano de criação, tipo, software, recuperação e tipo de acesso

Fonte: Elaborado pelas autoras

O mais antigo dos repositórios analisados é o da Universidad de Antioquia, criado em 2008, enquanto os outros repositórios foram implantados somente entre os anos de 2016 e 2022. Destes repositórios, a grande maioria

corresponde a repositórios institucionais acadêmicos que disponibilizam acesso a diferentes tipologias documentárias provenientes da produção acadêmica e de pesquisa das universidades. Entre esses documentos podem conter dados de pesquisas, como é o caso dos repositórios: Producción Académica UCC; SEDICI; Repositório Institucional UNESP; REPOUAM; e o Portal de Datos Abiertos da Universidad Nacional Autónoma de México (UNAM). Apenas seis repositórios identificados no re3data são específicos para dados de pesquisa.

Comparativamente, destaca-se o estudo realizado por Silveira (2021), onde foram analisadas as políticas de 18 repositórios universitários, em sua maioria dos Estados Unidos, e que disponibilizam as políticas em seus websites. Em comparação ao presente estudo, os repositórios da referida pesquisa foram criados na primeira década dos anos 2000, correspondendo a um período maior de consolidação de práticas de armazenamento, preservação e disponibilização de conteúdo digital. Isto posto, é possível inferir que a maturidade dos repositórios se impõe na apresentação de informações e normativas relacionadas aos mesmos e publicizadas em seus sites. Tal constatação pode explicar em parte a dificuldade em localizar informações no site ou repositórios latino-americanos.

No que se refere ao software utilizado na implementação dos repositórios, verifica-se que o DSpace está presente em 12 dos 20 repositórios analisados, possivelmente pelo fato de o DSpace ser um software livre, originalmente desenvolvido para a criação de repositórios e ser facilmente adaptado pelas instituições (IBICT, 2012). Em segundo lugar aparece o Dataverse, presente em quatro repositórios, sendo estes específicos para dados de pesquisa, podendo sua escolha ser justificada por ser um software desenvolvido para atender as demandas desse tipo de repositório (Pavão et al., 2018, p. 42). Em contrapartida, a Base de Dados Científicos da Universidade Federal do Paraná, utiliza-se do DSpace para a gestão de suas coleções de dados de pesquisa. Essa escolha provavelmente está atrelada ao uso do DSpace no repositório

institucional da Universidade. De acordo com Treloar, Groenewegen e Harboe-ree (2007), embora a abordagem de um único repositório para toda produção científica de uma instituição possa parecer atraente, existe uma série de desafios relacionados a curadoria da ampla variedade de documentos. Em estudo comparativo entre os sistemas DSpace e Dataverse, realizado por Rocha et al (2021), os autores relatam que um diferencial do Dataverse em relação ao DSpace, é que este apresenta recursos de configurações que possibilitam a criação de vários tipos de ambientes de repositório, enquanto o DSpace exige um esforço maior em suas configurações para se alcançar o mesmo resultado.

Quanto a verificação da disponibilização dos dados de pesquisa, pôde-se constatar que em sua maioria os repositórios têm por prerrogativa o acesso aberto, contudo permitem o acesso restrito aos conteúdos se assim for necessário. Em relação às licenças de uso dos dados de pesquisa, verifica-se que, na grande maioria dos repositórios a indicação é de licenciamento com as licenças Creative Commons, atribuição CC0 ou uma Atribuição-Não-Comercial sem obra derivada (CC BY-NC-ND 4.0).

## **5 Considerações finais**

Uma das limitações desta pesquisa refere-se à falta de informações nos repositórios e sites das bibliotecas e instituições para realizar uma pesquisa documental consistente. As políticas dos repositórios muitas vezes não fazem qualquer menção a dados de pesquisa, mesmo nos repositórios registrados na ferramenta re3data. Salienta-se a importância da publicização das políticas de repositórios, uma vez que estas são instrumentos oficiais norteadores das condutas a serem adotadas no acesso, depósito e uso do conteúdo armazenado nesses sistemas. Outro fato interessante e inesperado é a falta de dados de pesquisa nesses repositórios, constatando-se que o re3data não é totalmente confiável, possivelmente pela iniciativa de registro partir da própria



instituição, em uma autodeclaração. Sendo assim, infere-se que as informações registradas pelas universidades estejam em consonância com suas políticas institucionais, mesmo que não se tenha obtido sucesso nas buscas realizadas em alguns repositórios. Da mesma forma, como a maioria dos repositórios identificados são institucionais e não específicos de dados de pesquisa, infere-se que as buscas infrutíferas possam ser em decorrência de um processo de implementação de armazenamento desse conteúdo que ainda esteja em desenvolvimento nessas universidades.

Em relação a identificar a disponibilização de dados de pesquisa, como a maioria dos repositórios, 14 dos 20 investigados, não são exclusivos de dados de pesquisa, essa análise não foi satisfatoriamente possível. Dessa forma, considera-se que o objetivo do presente estudo não foi totalmente alcançado. A escolha por analisar repositórios universitários sem delimitar para repositórios específicos de dados de pesquisa acarretou prejuízo ao estudo. Da mesma forma, a utilização do re3data como uma única ferramenta de identificação dos repositórios parece não abranger o real cenário de iniciativas de implantação de repositórios de dados de pesquisa, a exemplo do Brasil.

Metodologicamente decidiu-se por analisar somente repositórios universitários, contudo no re3data recuperou-se um total de 39 outros repositórios que não fizeram parte do corpus desta pesquisa. Sendo assim, sugere-se para estudos futuros a análise desses repositórios para conhecer como se dá as iniciativas de abertura de dados de pesquisa nesses sistemas de informação.

## **Bibliografia**

- AMARO, B. (2021). Ciência aberta: desafios e perspectivas.  
<[https://www.youtube.com/watch?v=7r\\_n\\_nrMwts&t=6558s](https://www.youtube.com/watch?v=7r_n_nrMwts&t=6558s)>
- BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA. (2018a). Contenido.  
<<https://repositorioinstitucional.buap.mx/page/content>>

- BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA. (2018b). What are we?  
<<https://repositorioinstitucional.buap.mx/page/buapAbout>>
- BICKNESE, D. (2003-2004). Institutional repositories and the institution's repository: what is the role of university archives with an institution's on-line digital repository? *Archival Issues*, 28(2), 81-94.  
<[https://minds.wisconsin.edu/bitstream/handle/1793/45952/MA28\\_2\\_2.pdf?sequence=3&isAllowed=y](https://minds.wisconsin.edu/bitstream/handle/1793/45952/MA28_2_2.pdf?sequence=3&isAllowed=y)>
- ESTEVÃO, J. S. B. (2018). Apresentação da Base de Dados Científicos da Universidade Federal do Paraná. Curitiba. 50 slides.  
<[https://acervodigital.ufpr.br/bitstream/handle/1884/56095/Apresenta%3a7%c3%a3o\\_BDC%20%20UFPR.pdf?sequence=1&isAllowed=y](https://acervodigital.ufpr.br/bitstream/handle/1884/56095/Apresenta%3a7%c3%a3o_BDC%20%20UFPR.pdf?sequence=1&isAllowed=y)>
- IBICT. (2012). Sistema para Construção de Repositórios Institucionais Digitais (DSpace): sobre o DSpace.  
<<http://sitehistorico.ibict.br/pesquisa-desenvolvimento-tecnologico-e-inovacao/Sistema-para-Construcao-de-Repositorios-Institucionais-Digitais>>
- MACGREGOR, G. (2023). Digital repositories and discoverability: definitions and typology. In: Discoverability in digital repositories: systems, perspectives, and user studies.  
<[https://strathprints.strath.ac.uk/78405/7/Macgregor\\_2022\\_Digital\\_repositories\\_and\\_discoverability\\_definitions\\_and\\_typology.pdf](https://strathprints.strath.ac.uk/78405/7/Macgregor_2022_Digital_repositories_and_discoverability_definitions_and_typology.pdf)>
- PAVÃO, C. et al. (2023). Planejamento e incubação de repositórios de dados em ICTs brasileiras. In: Actas de la XI Conferencia Internacional de Bibliotecas y Repositorios Digitales, (p. 233-249). La Plata: Universidad Nacional de La Plata: SEDICI.  
<<https://www.lume.ufrgs.br/bitstream/handle/10183/255620/001162849.pdf?sequence=1>>
- PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO. (2022). Sobre [laboratório LAMBDA]. <<https://www.maxwell.vrac.puc-rio.br/sobre.php>>
- ROCHA, R. P. DA ET AL. (2018). Acesso aberto a dados de pesquisa no Brasil: soluções tecnológicas: relatório 2018.  
<<http://hdl.handle.net/10183/185126>>
- ROCHA, R. P. DA ET AL. (2021). Análise dos sistemas DSpace e Dataverse para repositórios de dados de pesquisa com acesso aberto. *Revista Brasileira de Biblioteconomia e Documentação*, 17, 1-25.  
<<https://brapci.inf.br/index.php/res/download/160968>>
- SALES, L. F. E SAYÃO, L. F. (2019). Uma proposta de taxonomia para dados de pesquisa. *Conhecimento em Ação, Rio de Janeiro*, 4(1), 31-48.  
<<https://revistas.ufrj.br/index.php/rca/article/download/26337/14573>>
- SAYÃO, L. F. E SALES, L. F. (2016). Algumas considerações sobre os repositórios digitais de dados de pesquisa. *Informação & Informação*, 21(2), 90-115.

<https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/27939/20122>

- SILVEIRA, J. I. E VANZ, S. A. de S. (no prelo). Diretrizes para políticas de depósito, acesso e uso de dados de pesquisa: proposta a partir da análise de repositórios de dados universitários internacionais. AtoZ Novas Práticas em Informação e Conhecimento.
- TORINO, E., ROA-MARTÍNEZ, S. M. E VIDOTTI, S. A. B. G. (2020). Dados de pesquisa: disponibilização ou publicação? In: Shintaku, M.; Sales, L. F e Costa, M. (Orgs). Tópicos sobre dados abertos para editores científicos p. 183-201. Botucatu, SP: ABEC, <[doi:10.21452/ 978-85-93910-04-3.cap15](https://doi.org/10.21452/978-85-93910-04-3.cap15)>
- TRELOAR, A., GROENEWEGGEN, D., E HARBOE-REE, C. (2007). The Data Curation Continuum: Managing Data Objects in Institutional Repositories. *D-Lib Magazine*, 13(9-10). <<http://www.dlib.org/dlib/september07/treloar/09treloar.html>>
- UNIVERSIDAD CATÓLICA DE CÓRDOBA. (2021). Política del Repositorio Institucional Producción Académica. <<http://pa.bibdigital.ucc.edu.ar/policias.html>>
- UNIVERSIDAD CATÓLICA DE SALTA. (2017). Resolución Rectoral nº 1017/17. <[http://bibliotecas.ucasal.edu.ar/opac\\_css/doc\\_num\\_data.php?explnum\\_id=1084](http://bibliotecas.ucasal.edu.ar/opac_css/doc_num_data.php?explnum_id=1084)>
- UNIVERSIDADE ESTADUAL PAULISTA. (2022). Apresentação. <<https://repositorio.unesp.br/>>
- UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS. (2022). Formato Licencia de uso y publicación. <<https://repository.udistrital.edu.co/handle/11349/3841>>
- UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO. (2022). Sobre el Portal de Datos Abiertos UNAM. <<https://datosabiertos.unam.mx/informacion/sobreportal.html>>
- UNIVERSIDAD NACIONAL DE LA PLATA. (2022). Políticas del repositorio. <<http://sedici.unlp.edu.ar/pages/politicas>>
- UNIVERSIDAD NACIONAL DE ROSARIO. (2022). Se presentó el Repositorio de Datos Académicos. <<https://unr.edu.ar/la-unr-presento-su-repositorio-de-datos-academicos/>>
- VANZ, S. A. DE S. E SILVA FILHO, R. C. (2019). O protagonismo das revistas na comunicação científica: histórico e evolução. In: Carneiro, F. F. B., Ferreira Neto, A. e Santos, W. dos. (Org.). *A comunicação científica em periódicos*. (1ed.), p. 19-44. Curitiba: Appris.

# El rol de la biblioteca universitaria en la gestión de datos de investigación: una estrategia para potenciar la visibilidad científica y la ciencia abierta

**DIANA MARÍA AYALA VILLADA**

Universidad Cooperativa de Colombia

[dianam.ayala@ucc.edu.co](mailto:dianam.ayala@ucc.edu.co)

## **EJE TEMÁTICO**

Datos abiertos

## **RESUMEN**

La gestión de datos de investigación se ha convertido en un aspecto fundamental de la actividad investigativa en el contexto del acceso abierto. La biblioteca universitaria, como institución clave en la gestión de información científica, desempeña un papel fundamental en la promoción y apoyo a la gestión de datos de investigación como una estrategia para potenciar la visibilidad científica y fomentar la ciencia abierta. En este artículo aborda el papel de la biblioteca universitaria en la gestión de datos de investigación, se describen las principales funciones y servicios que puede ofrecer en este ámbito, y se destacan las ventajas de integrar la biblioteca universitaria como aliado estratégico en la implementación de una política de gestión de datos. Específicamente se muestra la experiencia de la Universidad Cooperativa de Colombia, en el proceso de implementación de su estrategia de gestión de datos de investigación, su articulación con otras áreas y su contribución a potenciar la visibilidad científica y a promover la ciencia abierta, así como los nuevos retos y oportunidades que surgen para los profesionales de la información.

## **PALABRAS CLAVE**

Biblioteca universitaria; gestión de datos de investigación; visibilidad científica; ciencia abierta. University library; research data management; scientific visibility; open science.

## Introducción

La ciencia abierta y el acceso abierto a los resultados de investigación se han convertido en principios fundamentales en el panorama científico actual. La gestión de datos de investigación, entendida como el conjunto de procesos y prácticas que permiten organizar, describir, preservar y compartir los datos generados durante la actividad investigativa, juega un papel crucial en la promoción de la ciencia abierta y la visibilidad científica. En este contexto, la biblioteca universitaria se posiciona como una institución clave en la gestión de datos de investigación, ofreciendo servicios y recursos para apoyar a los investigadores en esta tarea.

La biblioteca universitaria cuenta con una larga trayectoria en la gestión de información científica, brindando servicios de acceso a la literatura científica, gestión de repositorios institucionales y apoyo en la difusión de la producción científica de la institución. Sin embargo, en el marco de la ciencia abierta, la gestión de datos de investigación se ha convertido en una necesidad creciente y compleja. Los datos de investigación generados por los investigadores son valiosos activos científicos, y su gestión adecuada puede contribuir a potenciar la visibilidad y el impacto de la investigación, así como a fomentar la colaboración y la reutilización de datos.

En este artículo se aborda el papel de la biblioteca universitaria en la gestión de datos de investigación como una estrategia para potenciar la visibilidad científica y la ciencia abierta. Se describirán las principales funciones y servicios que puede ofrecer la biblioteca universitaria en este ámbito, destacando su rol como aliado estratégico para los investigadores en la implementación de políticas y prácticas de gestión de datos de investigación, identificando los retos y oportunidades que representa para las bibliotecas universitarias en la actualidad.

## Gestión de datos de investigación y la ciencia abierta

La ciencia lo largo de la historia ha pasado por varias etapas; en sus inicios era empírica y su objeto de estudio eran los fenómenos naturales. Posteriormente pasó a ser teórica, incorporando el uso de modelos y generalizaciones. En tiempos más recientes se incorpora toda la revolución de la informática y se abre paso a la simulación y análisis más complejos de los fenómenos. Actualmente se habla ya del cuarto paradigma de la ciencia que implica la exploración de datos: la eScience.

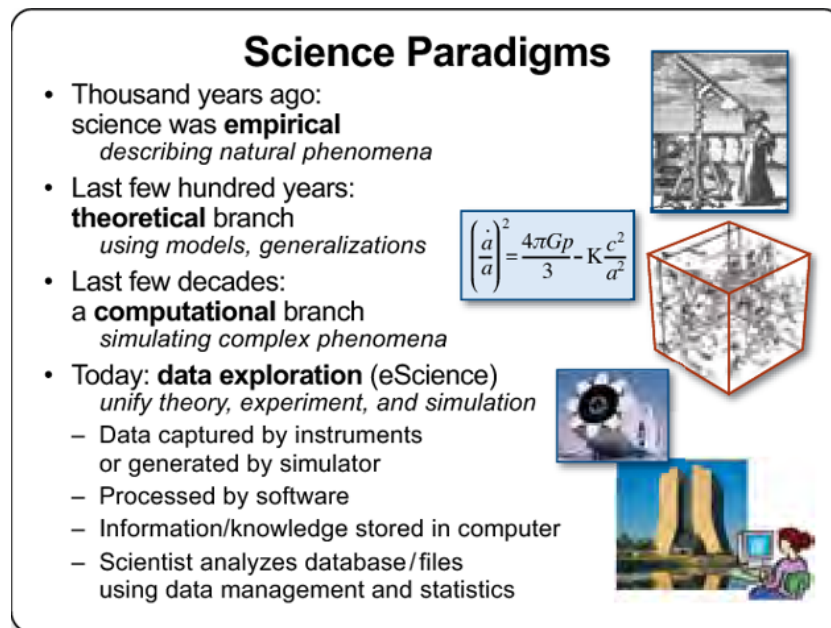


FIGURA 1. Paradigmas de la ciencia

Fuente: Jim Gray on eScience: A Transformed Scientific Method  
The Fourth Paradigm, Jim Gray, págs. XXVII-XXXI

Hoy la ciencia, enfrenta en sus diversas disciplinas, un cambio enorme: manejar una desbordada cantidad de datos, almacenarlos, procesarlos, hacer curaduría, analizarlos y difundirlos. En este sentido Jim Gray (2009) “aboga por la creación de nuevas herramientas que apoyen todo el ciclo de la investigación, desde la captura de datos, la curación y análisis hasta la

visualización de los mismos. Los datos que finalmente se publican son sólo la punta del iceberg”. Es en este punto donde la gestión de datos de investigación cobra total relevancia y vigencia y se convierte en una gran oportunidad para las bibliotecas universitarias, en tanto que tienen todas las fortalezas para liderar esta gestión desde un enfoque sistemático y para administrar los datos a lo largo de todo su ciclo de vida, desde la planificación inicial hasta el depósito final.

La CEPAL (2020) define los datos de investigación como los “datos que son recolectados, observados o creados para ser analizados y producir resultados de investigación originales”. Estos datos constituyen la información recopilada y utilizada en un estudio o investigación científica para responder a una pregunta de investigación o probar una hipótesis. Pueden ser de diferentes tipos: cuantitativos (números, mediciones) o datos cualitativos (textos, entrevistas). Los datos de investigación se obtienen a través de diversos métodos, como experimentos, encuestas, observaciones, análisis de documentos o recopilación de datos existentes.

La gestión de datos de investigación puede implicar la recopilación directa de nuevos datos específicamente para el estudio (por ejemplo, a través de encuestas o experimentos) o la utilización de datos ya existentes (por ejemplo, en bases de datos, registros o estudios anteriores). Es importante garantizar que los datos de investigación sean confiables, válidos y estén bien documentados para respaldar la interpretación y replicación de los resultados. Una vez recopilados, los datos de investigación se someten a un análisis estadístico o cualitativo, según corresponda, para extraer conclusiones y obtener resultados que respondan a la pregunta de investigación planteada.

En esta misma línea y en concordancia con los principios de la ciencia abierta se encuentran los datos de investigación abiertos que se refieren a datos generados o recopilados durante una investigación científica que están disponibles para el acceso, la reutilización y el intercambio público sin restricciones o barreras de acceso. Estos datos se comparten en formatos

legibles por máquina y suelen ser de libre uso, redistribución y modificación, siempre y cuando se atribuya su origen y se respeten posibles restricciones legales o éticas.

Los datos de investigación abiertos pueden abarcar diversas áreas del conocimiento, como ciencias naturales, ciencias sociales, ciencias de la salud, ingeniería, humanidades, entre otros. Algunos ejemplos de datos de investigación abiertos incluyen conjuntos de datos científicos, encuestas, registros de observaciones, datos genómicos, datos climáticos, datos geoespaciales, entre otros.

La apertura de los datos de investigación abiertos se ha convertido en una práctica cada vez más valorada y promovida en la comunidad científica y en instituciones académicas y gubernamentales. La disponibilidad de estos datos contribuye a la reproducibilidad de los estudios, permite una mayor colaboración entre investigadores, facilita la innovación y beneficia a la sociedad en general al permitir el acceso a información valiosa para la toma de decisiones informadas y el desarrollo de soluciones basadas en evidencia.

Es así como la gestión de datos de investigación y la ciencia abierta están estrechamente relacionados ya que la gestión adecuada de los datos de investigación es una parte fundamental de la ciencia abierta. En este sentido la ciencia abierta es un movimiento que promueve el acceso libre y sin restricciones a la investigación y a sus datos, lo que permite a cualquier persona reproducir, replicar y reutilizar los resultados de la investigación y es ahí donde la gestión adecuada de los datos de investigación es esencial para lograr estos objetivos.

La gestión de datos de investigación implica el proceso de planificación, organización, almacenamiento, documentación, preservación y uso de los datos de investigación. La buena gestión de los datos de investigación asegura que los datos sean confiables, accesibles, comprensibles y reutilizables a largo plazo.



Al hacer que los datos de investigación estén disponibles para la comunidad científica y para el público en general, se fomenta la transparencia y la rendición de cuentas en la investigación científica. Además, la apertura de los datos de investigación también puede aumentar la eficiencia de la investigación, ya que permite a otros investigadores construir sobre los resultados existentes y evitar duplicar esfuerzos.

Como dicen Barandiaran y Vila-Viñas, (2015), “implica la necesidad de analizar la ciencia abierta no como una nueva creación o tipo de ciencia separada del sistema ya existente”, sino la creación de una nueva forma de liberación de los insumos y resultados de la ciencia que se había ido cerrando y privatizando de manera progresiva; es decir, “no es una nueva ciencia, es la forma de hacer ciencia el siglo XXI” (Watson, 2015).

### **Gestión de datos de investigación y la visibilidad científica**

La gestión de datos de investigación es un elemento clave para mejorar la visibilidad científica de los investigadores y de las universidades. Al gestionar los datos de investigación de manera adecuada y ponerlos a disposición en repositorios públicos, se incrementa la accesibilidad de los datos y se facilita su uso y reutilización por parte de otros investigadores, esto a su vez aumenta la visibilidad de los resultados de investigación y el impacto de la producción científica. Al generarse un DOI (Digital Object Identifier) se impulsan las citaciones por cada set de datos publicados, ampliando mucho más las posibilidades, ya que tradicionalmente las citaciones solo se recibían por la publicación de los productos finales.

De igual manera la reutilización de datos de investigación puede generar nuevas investigaciones y descubrimientos, lo que puede contribuir a mejorar la visibilidad científica y a crear nuevas comunidades de colaboración académica a nivel nacional e internacional. También se aumenta la transparencia de los resultados de investigación y puede ayudar a mejorar la reputación y

credibilidad de los investigadores y las instituciones, a la vez que se está fomentando la ciencia ciudadana y la apropiación social del conocimiento.

### **Funciones y servicios de la biblioteca universitaria en la gestión de datos de investigación**

La ALA (American Library Association) (1990) define la biblioteca universitaria como “una combinación orgánica de personal, colecciones e instalaciones cuyo propósito es ayudar a sus usuarios en el proceso de transformar la información en conocimiento”. En este sentido tiene la biblioteca universitaria enormes retos, más aún en el momento actual donde la cantidad de información es cada vez mayor y el acceso a la misma crece exponencialmente, tanto que es difícil determinar su veracidad, calidad e impacto. “Para seguir siendo competitivos en un panorama científico cada vez más digitalizado y abierto, las bibliotecas académicas deberían monitorear permanentemente el contexto y desarrollar ofertas orientadas a la demanda a lo largo del ciclo de investigación” (Ogungbeni et al., 2016).

En consecuencia, es fundamental que la biblioteca universitaria se encuentre inserta dentro de un programa integral al servicio de la investigación institucional, incluso en un panorama ideal podría además liderarlo. Un claro ejemplo de esto puede ser la experiencia de los RIM:

Research information management (RIM) is the aggregation, curation, and utilization of information about research and is emerging as an area of increasing interest and relevance in many university libraries. RIM intersects with many aspects of traditional library services in discovery, acquisition, dissemination, and analysis of scholarly activities, and does so through the nexus with institutional data systems, faculty workflows, and institutional partners. RIM adoption offers libraries new opportunities to support institutional and researcher goals (Bryant et al., 2017).

En el siguiente gráfico se observan los elementos que puede aportar un sistema para la investigación con la adecuada estructuración dentro de un sistema universitario:

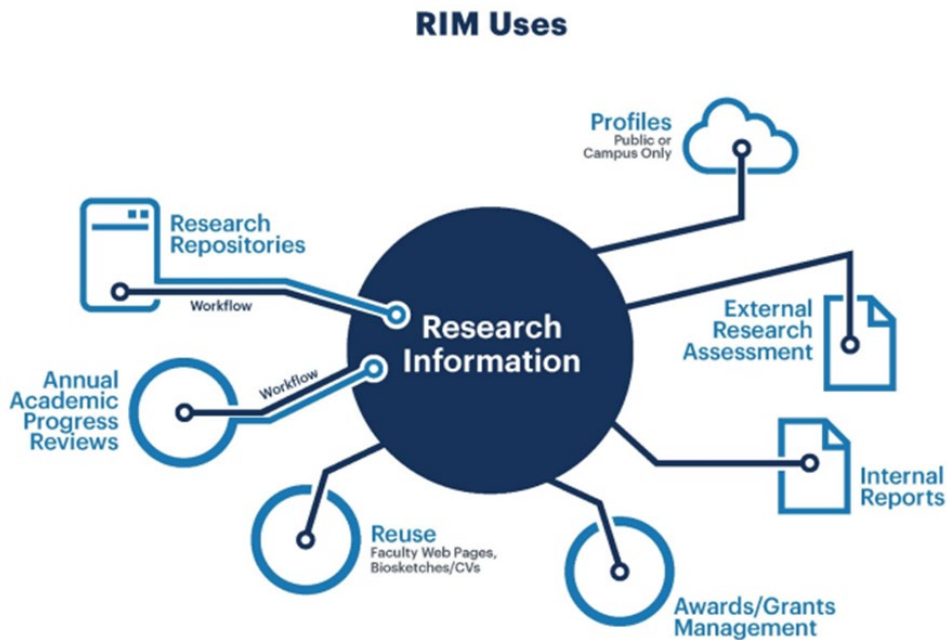


FIGURA 2. Usos del sistema RIM

Fuente: Bryant et al. (2017)

La gestión de la biblioteca universitaria desde su planeación tendrá cuatro ejes fundamentales de actuación que marcarán su línea de trabajo:

- La naturaleza de los datos.
- El creciente valor que la comunidad científica le otorga los datos
- El crecimiento exponencial de la cantidad de datos generados
- El uso de repositorios de datos para apoyar el ciclo de vida de las investigaciones.

Para que la biblioteca universitaria realmente pueda responder a este nuevo reto debe integrarse a la gestión de datos abarcando todo su ciclo vital comenzando desde la misma etapa de planificación de la investigación,

ejecución, análisis de resultados y publicación. Todo esto incluye en paralelo la creación de los conjuntos de datos de forma que estos sean precisos, completos, accesibles y reutilizables a lo largo del tiempo. Para tal efecto es fundamental que la biblioteca universitaria ofrezca servicios especializados y a la medida para el público investigador.



FIGURA 3. Gestión de datos de investigación en bibliotecas académicas

Fuente: Silva de Carvalho (2020)

Entre los principales servicios que debe enfocarse encuentran:

- Asesoría en la elaboración de planes de gestión de datos, esto implica la identificación de los tipos de datos generados, la definición de las políticas de retención, preservación y acceso, y la selección de herramientas y recursos adecuados para la gestión de los datos.
- Creación y gestión de repositorios de datos para el depósito y la preservación de los datos de investigación generados por los investigadores de la Universidad. Estos repositorios deben, en la medida de lo posible, promover acceso abierto y proporcionar una plataforma

para la difusión y visibilidad de los datos, siguiendo estándares internacionales para su indexación.

- Capacitación y formación en buenas prácticas de gestión de datos, la creación de planes de gestión de datos, la organización y documentación de los datos, y el uso de herramientas y recursos para la gestión de datos.
- Apoyo en la difusión y promoción de los datos de investigación, buenas prácticas de citación, la asignación de identificadores persistentes a los datos, y la promoción de los datos en repositorios institucionales y en otros repositorios de acceso abierto.

### **El bibliotecario de datos**

Ante estas nuevas dinámicas surge un nuevo rol dentro de las bibliotecas universitarias: el bibliotecario de datos, entendido como “personas entrenadas y especializadas en la curación, preservación y archivos de datos” (Ohaji, 2019). Su objetivo principal es garantizar que los datos estén organizados, accesibles, actualizados y seguros para su uso por parte de los usuarios autorizados.

Las funciones de un bibliotecario de datos pueden variar dependiendo del contexto, pero generalmente incluyen:

- Recopilación y adquisición de datos: El bibliotecario de datos es responsable de identificar las fuentes de datos relevantes y recopilarlos de manera sistemática. Esto puede implicar la búsqueda de datos internos en diferentes sistemas de la organización o la adquisición de datos externos de proveedores o fuentes públicas.
- Organización y catalogación: Una vez que los datos son recopilados, el bibliotecario de datos los organiza y cataloga de manera estructurada. Esto implica asignar etiquetas, metadatos y clasificaciones a los datos para facilitar su búsqueda y recuperación posterior.

- **Mantenimiento y actualización:** Los datos requieren un mantenimiento regular para asegurar su integridad y calidad. El bibliotecario de datos se encarga de aplicar actualizaciones, correcciones y limpieza de datos, así como de supervisar cualquier cambio en las fuentes de datos para asegurar que se mantengan precisos y actualizados.
- **Control de acceso y seguridad:** La gestión de la seguridad de los datos es fundamental para proteger la información sensible y garantizar el cumplimiento de las regulaciones y políticas internas. El bibliotecario de datos establece y aplica los controles de acceso apropiados, asegurándose de que solo los usuarios autorizados puedan acceder a los datos.
- **Soporte y asistencia:** El bibliotecario de datos puede brindar soporte y asistencia a los usuarios que necesitan acceder a los datos. Esto puede implicar responder preguntas, proporcionar orientación sobre la interpretación de los datos o ayudar en la resolución de problemas técnicos relacionados con el acceso a los datos.
- **Gestión del ciclo de vida de los datos:** Los datos tienen un ciclo de vida que abarca desde su creación o adquisición hasta su eliminación o archivo. El bibliotecario de datos se encarga de gestionar este ciclo, asegurándose de que los datos sean adecuadamente respaldados, archivados y eliminados según las políticas establecidas.

Estas nuevas funciones se convierten en retos para la profesión y lógicamente implican una preparación diferencial y exhaustiva tal como lo menciona Spinak (2019):

### Competencias del *data librarian*:

- Datos y tecnologías geoespaciales;
- Lenguajes de programación;
- Paquetes estadísticos (SPSS, Stata, SAS);
- Gestión y preservación de datos;
- Redacción de solicitudes de fondos de investigación;
- Gestión de proyectos;
- Pensamiento crítico y resolución de problemas;
- Empaquetado de datos, reutilización de datos, codificación de datos;
- Repositorios institucionales y metadatos;
- Tecnologías actuales online;
- Análisis de datos cualitativos y cuantitativos.

### **Papyrus: estrategia para la gestión de datos de la Universidad Cooperativa de Colombia**

Papyrus, es una iniciativa que busca contribuir al movimiento de ciencia abierta en Colombia, y surge de la necesidad de 60 instituciones de educación superior y algunos centros de investigación, miembros del Consorcio Colombia, para darle visibilidad a la producción intelectual y a los datos de investigación producto de ésta; facilitando el ciclo de reutilización de estos como punto de partida para incentivar nuevas producciones basadas en un soporte confiable y posibilitando a su vez, minimizar costos en las investigaciones e incentivar el trabajo colaborativo bajo la filosofía de ciencia abierta.

Inicia en el año 2022, con 13 instituciones de educación superior (Universidad de los Andes, Colegio mayor nuestra señora del Rosario, Universidad Pontificia Bolivariana, Universidad de Caldas, Universidad de Cartagena, Universidad de la Costa, Universidad de Medellín, Universidad El Bosque, Universidad Santo Tomás, Universidad EAN, Fundación Universitaria del Área Andina, Universidad Católica Luis Amigó y Universidad Cooperativa de

Colombia) para desarrollar un servicio que permita la integración, organización y disposición de los datos producidos en estas instituciones resultado de los diferentes procesos de investigación.

En la Universidad Cooperativa de Colombia, específicamente, la gestión de información y el conocimiento han adquirido gran relevancia por el apoyo que brinda a los procesos académicos y de investigación en el ámbito de la producción intelectual y académica. De este modo el acceso a la información y a los datos de investigación se consolidan como una condición que cada vez toma más relevancia para validar y sustentar las publicaciones actualmente.

De igual forma, en la misión institucional, la Universidad tiene entre sus propósitos contribuir a la construcción y divulgación del conocimiento desde su quehacer académico y de investigación. Es así como desde el año 2014 se inicia en la Universidad el trabajo en diferentes frentes alrededor del acceso abierto, donde se resalta que en el año 2017 se dispuso el servicio de Repositorio institucional el cual permite compartir el conocimiento generado por la comunidad académica en el contexto mundial, y que es ratificado en el año 2018 con la creación de la Política de acceso abierto (Acuerdo superior N°430 de 2018).

## Iniciativas institucionales de Acceso abierto



FIGURA 4. Iniciativas institucionales de Acceso abierto

Fuente: Subdirección, Sistema de Información Bibliográfico UCC, (2022)



Papyrus, en la Universidad Cooperativa de Colombia, aborda todo el ciclo del dato de investigación articulando el trabajo de diferentes áreas de la universidad para ello. El proyecto es liderado desde el sistema de bibliotecas y se desarrolla en tres momentos: un primer momento de planeación que comprendió el diseño del proceso, la estructura de funcionamiento, la documentación de los procedimientos, la actualización y creación de las políticas, elaboración de guías y demás materiales que se requieran para dejar establecido como proceso dentro de la Institución.

Un segundo momento de implementación, inició con la configuración en la plataforma Dataverse del Repositorio de datos de investigación de la universidad y la gestión de todos los requerimientos técnicos que esto implica, teniendo en cuenta estándares internacionales, niveles de acceso y las medidas necesarias para garantizar la protección de datos personales establecidas por la ley. Al contar con el repositorio de datos de investigación se procedió a realizar una prueba piloto para verificar su funcionamiento y la efectividad del diseño planteado.

Por último y a partir de los resultados y aprendizajes de la prueba piloto, se desarrolló el plan de capacitaciones a los grupos de interés, se desplegó el plan de comunicaciones y se puso a disposición de la comunidad universitaria la gestión de los datos de investigación, como un nuevo servicio para visibilizar la investigación.



FIGURA 5. Estructura del servicio de gestión de datos de investigación en la Universidad Cooperativa de Colombia

Fuente: Creación propia (2022)

#### Resultados obtenidos hasta el momento:

- Conformación de equipo base interdisciplinario con formado por el Sistema de bibliotecas, Dirección de Innovación y Tecnología Educativas, Dirección de investigaciones, Dirección Editorial y Dirección de Transferencia del Conocimiento.
- Repositorio de datos funcionando y correctamente configurado.
- Guías y manuales publicados de manera oficial con el aval de la Dirección de Planeación y Efectividad.
- 70 dataverses creados, cada uno correspondiente a un proyecto de investigación.
- 167 datasets creados.
- 6 líderes formados y distribuidos por campus.
- Creación de un micrositio con recursos para el uso de Papyrus.

- Inclusión de la gestión de datos dentro de las convocatorias para proyectos de investigación dentro de la Universidad.
- Indexación del repositorio de datos en Google Data Search y Datacite Commons.

## Conclusiones

La gestión de datos de investigación es crucial en el proceso de investigación científica y tiene un impacto significativo en la calidad, eficiencia y credibilidad de la investigación. La gestión adecuada de los datos de investigación permite a otros investigadores comprender y replicar los resultados de investigación, esto contribuye a la reproducibilidad y transparencia en la investigación, lo que a su vez mejora la credibilidad y confianza en los resultados obtenidos.

La gestión de datos de investigación mejora la eficiencia del proceso de investigación; al gestionar adecuadamente los datos, se pueden reducir los costos y el tiempo requerido para recopilar, analizar y compartir los datos. Los datos de investigación pueden ser reutilizados para nuevos proyectos, se puede aumentar la posibilidad de que los datos sean útiles para futuros proyectos de investigación, lo que a su vez puede ahorrar tiempo y costos en la recopilación de nuevos datos.

Muchas instituciones, agencias de financiamiento y editoriales exigen el cumplimiento de políticas de gestión de datos de investigación, lo que además fomenta el acceso abierto. La gestión adecuada de los datos de investigación garantiza la preservación de los datos a largo plazo para futuras generaciones de investigadores, en plataformas seguras y de fácil acceso.

Para lograr un resultado exitoso en la gestión de datos de investigación es fundamental el trabajo interdisciplinario; la biblioteca debe crear alianzas estratégicas con el área de Investigaciones, la Editorial, Transferencia del Conocimiento y demás dependencias relacionadas a la temática. A la

Universidad Cooperativa de Colombia, la gestión de datos le ha permitido consolidar el trabajo Multicampus entre las 23 bibliotecas que conforman el sistema, donde los liderazgos regionales han sido un factor de éxito.

La biblioteca universitaria tiene una oportunidad enorme en el liderazgo de los procesos de gestión de datos de investigación; de ser ese eje articulador que propicie que las Universidades avancen cada vez hacia la ciencia abierta y a la vez refuerzan su rol estratégico dentro de la institución.

El trabajo en Red es clave para lograr un mayor impacto, en este caso puntual, los datos publicados en el Repositorio de la Universidad Cooperativa de Colombia quedan automáticamente conectados en Papyrus nacional, junto con las demás universidades, lo que genera nuevas comunidades, trabajo colaborativo y maximiza el impacto de los resultados.

## Bibliografía

- AMERICAN LIBRARY ASSOCIATION (1990). Glosario ALA de Bibliotecología y Ciencias de la Información.
- BARANDIARAN, X. E.; VILA-VIÑAS, D. (2015). Buen conocer/FLOK Society: modelos sostenibles y políticas públicas para una economía social del conocimiento común y abierto en Ecuador / (eds.).— 1.ª ed. — Quito: Editorial IAEN / Editorial CIESPAL.
- BRYANT, R., CLEMENTS, A., FELTES, C., GROENEWEGEN, D., HUGGARD, S., MERCER, H., MISSINGHAM, R., OXNAM, M., RAUH, A. & WRIGHT, J. (2017). Research Information Management: Defining RIM and the Library's Role. Dublin, OH: OCLC Research. DOI: <<https://doi.org/10.25333/C3NK88>>
- CEPAL (2020). Gestión de datos de investigación. <<https://biblioguias.cepal.org/gestion-de-datos-investigacion/tiposdatos#:~:text=%22Los%20datos%20de%20investigaci%C3%B3n%20son%20registros%20concretos%2C%20que%20pueden%20tomar,validaci%C3%B3n%20de%20hallazgos%20de%20investigaci%C3%B3n.%E2%80%9D>>
- OGUNGBENI, J. I. ET AL. (2018). The roles of academic libraries in propagating open science: a qualitative literature review. Information Development, 34(2). DOI: <<https://doi.org/10.1177/0266666916678444>>

- GRAY, J. (2009). The Fourth Data-Intensive scientific Discovery Paradigm. United States of America Microsoft, p. 5.
- OHAI, I. K., CHAWNER, B. & YOONG, P. (2019). The role of a data librarian in academic and research libraries. *Information Research*, 24(4), 844. <<https://web.archive.org/web/20191213195857/http://InformationR.net/ir/24-4/paper844.html>>
- SILVA DE CARVALHO, E. R., LIMA LEITE, F. C., ROCHA BELLO BERTIN, P. (2020). Bibliotecas académicas e gestão de dados de pesquisa: uma revisão bibliográfica. *Investigación Bibliotecológica*, 35(86), 99-121. DOI: <<https://doi.org/10.22201/iibi.24488321xe.2021.86.58266>>
- SPINAK, E. (2019). La especialidad Bibliotecarios de Datos en la Ciencia Abierta. SciELO en Perspectiva. <<https://blog.scielo.org/es/2019/11/01/la-especialidad-bibliotecarios-de-datos-en-la-ciencia-abierta/>>
- WATSON, M. (2015). When will 'open science' become simply 'science'?. *Genome Biology*, 16(1). <<http://genomebiology.biomedcentral.com/articles/10.1186/s13059-015-0669-2>>

# Criptografia de arquivo em Biblioteca Digital com o software Dspace

## **LUCAS ÂNGELO DA SILVEIRA**

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA  
UNIVERSIDADE FEDERAL DO RIO GRANDE

## **MIRELE CAROLINA SOUZ FERREIRA COSTA**

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA  
UNIVERSIDADE FEDERAL DO RIO GRANDE

## **MILTON SHINTAKU**

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA  
UNIVERSIDADE FEDERAL DO RIO GRANDE

## **INGRID TORRES SCHIESSL**

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA  
UNIVERSIDADE FEDERAL DO RIO GRANDE

## **RAÍSSA DA VEIGA DE MENÊSES**

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA  
UNIVERSIDADE FEDERAL DO RIO GRANDE

## **RAFAEL CHAPLIN SCHRAMM**

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA  
UNIVERSIDADE FEDERAL DO RIO GRANDE

### **EIXO TEMÁTICO**

Infraestructura tecnológica

## RESUMO

**Introdução:** Após a popularização dos documentos em formato digital, muitos órgãos governamentais têm utilizado o software DSpace para manutenção de suas bibliotecas digitais e conseqüentemente problemas de segurança são recorrentes, pois muitos possuem dados sensíveis que precisam ser preservados. O presente estudo tem como objetivo descrever a implementação de criptografia para proteger documentos digitais no software DSpace contra acesso não autorizado. **Metodologia:** A metodologia da informática na criação de inovações é utilizada seguindo um modelo genérico fracamente acoplado, que pode ser aproveitado em bibliotecas digitais utilizando o DSpace. A técnica chamada Advanced Encryption Standard (AES) é usada para implementar a criptografia por meio de uma biblioteca java de código aberto. **Resultados:** Os relatórios obtidos a partir de testes de validação em ambientes de desenvolvimento e produção, mostram que a criptografia aumenta a segurança dos dados contra acesso não autorizado sem prejudicar o desempenho do DSpace, podendo ser utilizada por órgãos governamentais e outras instituições que desejam proteger seus dados.

## PALAVRAS-CHAVE

Criptografia; biblioteca digital; sistema de informação; DSpace.

## 1 Introdução

No final do século XX, surgiu o movimento dos arquivos abertos, em que o sistema de informação de maior destaque era a biblioteca digital responsável pela divulgação da literatura cinza, especialmente teses e dissertações. No Brasil, no início do século XXI, os órgãos governamentais passaram a utilizar as bibliotecas digitais para disseminar sua memória técnica e dar maior transparência à gestão pública.

A primeira biblioteca digital do governo no Brasil foi implementada usando o software DSpace e foi chamada de Biblioteca Digital Jurídica (BDJur) cujo objetivo era divulgar as informações jurídicas geradas no tribunal (Basevi, 2005). Desde então, diversos outros órgãos governamentais criaram suas bibliotecas digitais em todas as esferas e poderes.

Uma pesquisa com Macêdo, Shintaku e Brito (2015) 13 bibliotecas digitais vinculadas diretamente com órgãos governamentais brasileiros, todas projetadas com o DSpace.

Atualmente, um estudo realizado pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), constatou que existem cerca de 50 repositórios implementados.

Ao contrário dos repositórios institucionais, as bibliotecas digitais governamentais publicam, em sua maioria, documentos únicos pertencentes à memória técnica, colocando-os como canais de publicação, mesmo que informais. Assim, a questão que envolve a segurança de acesso à informação deve ser considerada, pois alguns documentos digitais podem ser de acesso restrito, como é o caso de parte da memória técnica produzida pelas forças armadas de uma nação.

A restrição de acesso existe no DSpace desde sua primeira versão, na qual pode ser restringida em diversos níveis, sendo gerenciada através de login. Ser capaz de dar permissão de acesso a todos os itens de uma comunidade ou coleção, a um item específico ou a um arquivo dentro de um item para usuários ou grupos de usuários específicos. No entanto, todos os documentos depositados no DSpace estão em formato claro, ou seja, legíveis se acessados indevidamente. Assim, mesmo que o acesso aos arquivos pertencentes a um item seja restrito, isso não impede que sejam lidos em caso de eventual invasão, comprometendo o sigilo de informações confidenciais.

Este trabalho apresenta os resultados referentes à implementação de uma trava de segurança extra no *bitstream*, armazenado no DSpace para a biblioteca digital da Secretaria de Assuntos Estratégicos da Presidência da República (SAE) da Presidência da República.

Implementamos técnicas de criptografia de forma transparente para o usuário final em itens que possuem etiqueta confidencial no momento do depósito. Embora grande parte das implementações do DSpace sejam voltadas



para sistemas de informação vinculados ao movimento *Open Access*, existem documentos que requerem restrição, como os relacionados à geração de patentes ou mesmo aqueles que possuem certa sensibilidade quanto ao seu conteúdo. Nossos resultados têm aplicação prática na secretaria de assuntos estratégicos do Brasil, cujo repositório está publicado em [este link](#).

## 2 Criptografia

Técnica originária do antigo Egito, adaptada para segurança da informação para maior segurança na comunicação e prevenção de acessos indevidos. O autor Shannon (1949) considerado o pai da teoria da informação, foi um dos precursores da criptografia de base matemática e computacional, apresentando um esquema criptográfico Figura 1.

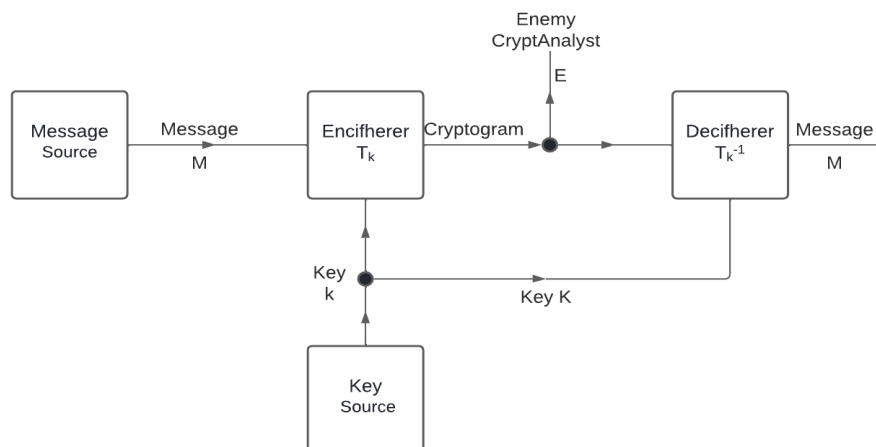


FIGURA 1. Esquema de criptografia

Fonte: Shannon (1949, p. 661)

A mensagem a ser criptografada passa por um criptografador gerando um criptograma, com o auxílio de uma chave. Se esse criptograma for acessado por alguém que não possua a chave de descryptografia, essa pessoa não terá acesso ao conteúdo de forma legível, apenas a um conjunto de bytes codificados que não fazem sentido a olho nu. Por outro lado, o destinatário, de

posse da chave de criptografia, consegue decodificar o criptograma e ter acesso à mensagem.

Os algoritmos de criptografia são classificados especificamente como simétricos e assimétricos. A criptografia simétrica utiliza uma única chave, que é compartilhada entre o remetente e o destinatário de um conteúdo. Esta chave é uma string de bits, que vai definir como o algoritmo irá criptografar um conteúdo (Aufa, Endroyono, & Affandi, 2018; Simmons, 1979)

A criptografia assimétrica é baseada em 2 tipos de chaves de segurança, uma privada e outra pública, e são usadas para criptografar mensagens e verificar a identidade de um usuário (Yassein, Aljawarneh, Qawasmeh, Mardini, & Khamayseh, 2017). A chave privada é usada para descriptografar as mensagens, enquanto a chave pública é usada para criptografar o conteúdo.

O estudo sistemático em Diffie e Hellman (2022) aborda os conceitos criptográficos disponíveis e identifica diferentes direções e problemas de pesquisa e que os profissionais escolhem um conceito adequado para seu caso de uso específico. Na pesquisa Abood e Guirguis (2018) são estudados os algoritmos de criptografia mais importantes da literatura, a fim de melhorar o desempenho dos métodos criptográficos. Em grande medida, os métodos de criptografia provaram ter suas vantagens e desvantagens e provaram ser adequados para diferentes aplicações. O estudo forneceu evidências suficientes para afirmar que o algoritmo mais confiável é o *Advanced Encryption Standard* (AES) em termos de velocidade de criptografia, descriptografia, complexidade, tamanho da chave, estrutura e flexibilidade.

### 3 Procedimentos metodológicos

O presente trabalho aplicou aspectos da área de informática alinhada com a ciência e gestão da informação em bibliotecas digitais implementadas com o software DSpace. O estudo possibilita que a criptografia dos arquivos

depositados ocorra de forma transparente e eficaz, de acordo com os objetivos traçados.

O estudo de caso apresentado possui aspectos exploratórios, a fim de dar maior familiaridade ao uso da criptografia em bibliotecas digitais implementadas com DSpace. Além disso, o estudo possui aspectos de apresentação do produto como sugerido por Wazlawick (2014), pois apresenta uma abordagem inovadora, que pode ser adotada em futuras versões do DSpace. Nossa abordagem fornece funcionalidade genérica que pode ser facilmente replicada em outras bibliotecas digitais implementadas com o DSpace. Cabe ressaltar que a implementação da criptografia no DSpace para garantir a confidencialidade de documentos sensíveis é uma inovação na segurança do armazenamento, que, segundo o conhecimento dos autores, ainda não foi explorada. Neste contexto, o estudo foi realizado em cinco etapas, a saber:

- 1) Levantamento de requisitos: caracterização do problema, levantamento das necessidades e desenho da solução a ser alcançada;
- 2) Prospecção de soluções: levantamento e avaliação de tecnologias que atendem ao levantamento de requisitos;
- 3) Implementação da solução selecionada: implementação da tecnologia prospectada, ajustando se necessário;
- 4) Testes e ajustes: avaliação da implementação por meio do uso sistemático, com adaptações na tecnologia quando necessário;
- 5) Apresentação dos resultados: registro e disseminação dos resultados.

Essas etapas se aproximam do ciclo contínuo de desenvolvimento de software amplamente difundido na ciência da computação e áreas afins. Possivelmente uma das diferenças diz respeito ao objetivo de gerar conhecimento na resolução de problemas. Portanto, o estudo aborda o que Björk (2007) defende sobre o que a ciência faz: gerar novos conhecimentos e melhorar as condições de vida. Da mesma forma, vai ao encontro do que relata o referido autor sobre as motivações da ciência na resolução de problemas.

## **4 Resultados**

Como forma de relatar os resultados deste estudo, o texto foi dividido de acordo com as fases planejadas, descrevendo-as sequencialmente seguindo o ciclo de estudos na resolução de problemas de informática, mas mantendo o foco nas questões de informação. Portanto, apresenta os resultados sequencialmente para cada etapa, onde a etapa anterior alimenta a próxima.

### **4.1 Levantamento de requisitos**

Em estudos voltados para a criação ou adequação de ferramentas informatizadas, o primeiro passo é o levantamento de requisitos (Angelis, Ferrari, Gnesi, & Polini, 2018; Yaseen & Farooq, 2018). O estudo visa aumentar a segurança e validade dos dados depositados por meio de documentação em bibliotecas digitais. As necessidades podem ser representadas na Figura 2, na qual documentos confidenciais de caráter restrito são depositados na biblioteca digital de forma criptografada, somente sendo descriptografados por usuários com permissão. Para indicar ao DSpace qual depósito deve ser criptografado, usamos os metadados `dc.right.access` presentes no formulário de depósito.

## CRIPTOGRAFIA EM BIBLIOTECA DIGITAL

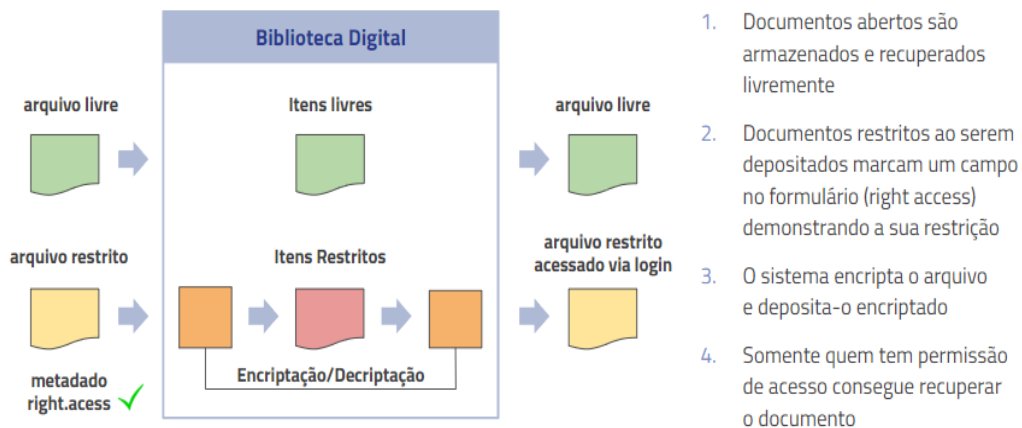


FIGURA 2. Fluxo da encriptação no DSpace

Fonte: Elaboração dos autores, 2023

Na arquitetura do serviço, a encriptação é transparente ao usuário ocorrendo de forma automatizada no momento em que um depósito ocorre. Para indicar quais depósitos envolvem documentos restritos, um novo campo do tipo checkbox é usado no formulário de entrada. Caso marcado como restrito, o encriptador é invocado e codifica o documento, armazenado-o de forma criptografada. Quanto à recuperação, os itens livres podem ser amplamente acessados pelos usuários anônimos. No entanto aos itens restritos, somente quem está logado e possui permissão consegue acessá-los. Os itens restritos armazenados de forma criptografada serão automaticamente decriptados para apresentação aos usuários com permissão.

### 4.2 Prospecção de solução

No decorrer do estudo chegou-se a duas opções a serem implementadas, sendo a primeira a integração de um software livre que atuaria como intermediário e a segunda implementando o encriptador/decriptador no DSpace. Nas duas opções levantadas o resultado seria o mesmo, podendo ser

identificada por integração ou adição, ficando totalmente transparente aos usuários.

Nesse sentido os resultados da prospecção revelou que o melhor caminho seria a adição de um pacote padrão Java, que implementa o modelo Advanced Encryption Standard (AES). Esta opção deu-se por apresentar o menor impacto a biblioteca digital, visto que adota um pacote na mesma tecnologia utilizada no DSpace, baseada em um padrão internacionalmente aceito em criptografia, adotada por, entre outros, o governo dos Estados Unidos da América.

O AES faz parte do conjunto de Application Programming Interface (API) chamada *Java Cryptography Extension*, propondo artefatos voltados à criptografia, entre os quais, voltados a implementação do algoritmo Rijndael, aprovado pelo National Institute of Standards and Technology (NIST), agência norte-americana voltada a promoção de inovação. O Modo Contador Galois, Criptografia GCM, é um algoritmo recomendado para criptografia autenticada com dados associados (Heron, 2009; Selent, 2010).

O GCM é construído a partir de uma cifra de bloco de chave simétrica com um tamanho de bloco podendo ser de: 128, 192 ou 256, tal que, quanto maior o bloco, mais complexo é a quebra da chave de decifração. Além disso, o GCM é um modo de operação do algoritmo AES conhecido na literatura como AES GCM o qual fornece garantia da autenticidade dos dados usando uma função de hash universal. Neste trabalho foi utilizada a biblioteca java bouncy castle, que implementa o AES GCM com tamanho de bloco de chaves de 256 bits, o que torna o depósito de dados confidenciais altamente confiável. Consulte a [documentação oficial](#) para obter mais detalhes sobre a biblioteca bouncy castle.

O modelo de encriptação AES foi criado em 1997 o que poderia estar suplantado por outros modelos (Paar & Pelzl, 2010). Entretanto, ao revisitar as implementações desse padrão, verificou que ainda se apresenta como atual, tanto que, mesmo com mais de 20 anos se apresenta de forma onipresente em

várias aplicações e possivelmente ainda continue sendo utilizada por muitas décadas seguintes (Menezes & Stebila, 2021).

### 4.3 Implementação da solução selecionada

A implementação da solução deu-se seguindo o padrão de desenvolvimento do DSpace, de forma a ter o menor dos impactos e possibilitar a portabilidade da solução para as novas versões do software. Da mesma forma, caso ocorra novas versões da solução implementada, pode-se atualizá-la com menor efeito no sistema, visto que, o DSpace versão 7, continua tendo compatibilidade no back-end com a versão 6. Na sequência, detalhes do serviço implementado são compartilhados.

#### 4.3.1 Elemento e qualificador no padrão Dublin Core

O Dublin Core (DC) é um padrão de metadados que por default é composto por 15 elementos e seus respectivos qualificadores usados para descrever recursos eletrônicos. O padrão DC possibilita refinar os elementos, dando maior especificidade. Por exemplo, o elemento *DC.contributor*, que especifica todos os usuários que atuaram na construção do documento, pode possuir um refinamento ainda mais preciso com *DC.contributor.author*, o qual é usado para especificar o autor do recurso a ser depositado. Vale ressaltar que, o padrão é flexível e genérico ao ponto de dar autonomia à criação de novos elementos/qualificador conforme a nossa necessidade para descrever um recurso eletrônico sempre que o leque de opções de elementos/qualificadores não nos atenderem.

No DSpace, o formulário de entrada, usado para compor os metadados de um depósito, na sua forma tradicional, adota o padrão DC. Dentre vários metadados necessários para descrever um novo recurso, vamos dar ênfase ao elemento *DC.rights*, usado para armazenar informações sobre os direitos sobre o recurso. Para o propósito do nosso trabalho, estendemos o grau de refinamento do elemento *DC.rights*, criando o qualificador *access*. Assim, o elemento qualificado novo *DC.rights.access* irá conter a informação se o

arquivo depositado será ou não criptografado. Esse processo de criação de um elemento ou qualificador novo no esquema de metadados Dublin Core é feito por meio do painel administrador no DSpace.

A inclusão de campos de metadados no formulário de entrada do DSpace é simples, por meio de um arquivo em formato XML chamado input-forms.xml. O trecho de código xml no formulário é apresentado na Figura 3 apresenta como foi codificado o input-forms.xml para incluir o novo campo de metadados para conter a informação se um arquivo será ou não encriptado.

```

<value-pairs      value-pairs-name="common_rights_access"      dc-
term="rights_access">
  <pair>
    <displayed-value>Acesso                Aberto</displayed-value>
    <stored-value>Acesso                    Aberto</stored-value>
  </pair>
  <pair>
    <displayed-value>Acesso                Restrito</displayed-value>
    <stored-value>Acesso                    Restrito</stored-value>
  </pair>
</value-pairs>

```

FIGURA 3. Código XML do input-forms.xml com a codificação de novo campo de metadados

Fonte: Elaboração dos autores, 2023

O resultado da inclusão do código no input-forms.xml no formulário de entrada é apresentado na Figura 4, como um campo de seleção do tipo checkbox. Neste tipo de campo, só é possível selecionar uma opção, por meio de um clique, no qual a informação contida na tag <stored-value> (acesso aberto ou acesso restrito) é inserida como o conteúdo do metadado *DC.right.access*.



Escolha entre acesso aberto ou acesso restrito.

Acesso à informação \*

Acesso Aberto  Acesso Restrito

FIGURA 4. Metadado *dc.rights.acess*

Fonte: Elaboração dos autores, 2023

Dessa forma, para iniciar o processo de criptografia, basta que o usuário marque o campo como Acesso restrito no *checkbox*, o qual ativa a função de criptografia implementada nos programas do DSpace que atua na carga do arquivo. Ao iniciar o processo de carga o DSpace verifica se esse metadado está ativo (flag) para invocar as subrotinas de criptografia que são descritas na sequência.

#### 4.3.2 Processo de submissão de um item

O diagrama de fluxo apresentado na Figura 5, reflete um depósito de ponta a ponta. Nesta seção daremos detalhes de como a ingestão de um novo depósito se comporta internamente no DSpace. Partindo do pressuposto de que, o usuário preencheu o formulário de descrição dos metadados, fez o upload do arquivo, o fluxo de carga tem os seguintes comportamentos. 1) se o metadado *DC.rights.access* está marcado como Acesso Aberto, o depósito possui fluxo normal e nada a ser feito com relação a encriptação. Enquanto que, *DC.rights.access* marcado como Acesso Restrito, o fluxo normal do DSpace sofre alteração. Alterações no fluxo precisam ser feitas no back-end do DSpace, explicitamente na classe *UploaStep.java*. Detalhes de implementação são omitidos, todavia, uma explicação em alto nível será dada.

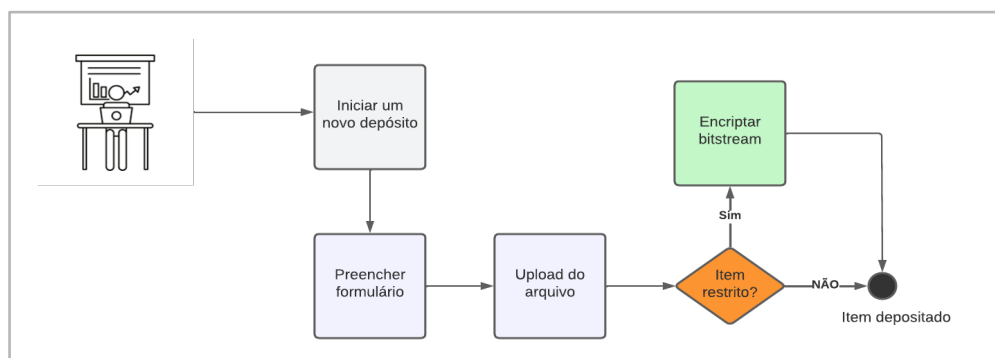


FIGURA 5. Diagrama de fluxo

Fonte: Elaboração dos autores, 2023

Inicialmente, usamos o valor do metadado *DC.rights.access* para desviar o fluxo, ou seja, quando marcado como “Acesso Restrito”. O *bitstream* (arquivo físico depositado) é manipulado, aplicando encriptação utilizando *Bouncy castle*, descrita na Seção 4.2 e posteriormente armazenado no diretório *assetstore*, o qual é usado pelo DSpace para armazenamento dos bitstreams. Detalhe, ao encriptar o arquivo físico, metadados não são afetados e nem os serviços disponibilizados pelo DSpace, ou seja, todos os serviços que funcionavam anteriormente, continuarão a funcionar posteriormente. Isso se deve ao fato de, o produto proposto neste trabalho ser fracamente acoplado, funcionando como um agregador de recursos.

#### 4.3.3 Processo de recuperação de um item

O processo de recuperação de um item depositado é descrito no diagrama de fluxo da Figura 6. Ao processar uma requisição de busca simples ou avançada, o DSpace irá verificar o valor do metadado *DC.rigth.access*, se o valor armazenado equivale a “Acesso Aberto”, então o fluxo padrão implementado no software será executado e o bitstream será disponibilizado ao usuário. Em contrapartida, *DC.rigth.access* com valor “Acesso Restrito” desencadeará uma série de passos extras de verificação e manipulação da informação.

A customização em caso de acesso restrito, foi realizada a nível de back-end na classe java chamada *itemTag.java*, o qual é responsável por tratar as informações antes de disponibilizar ao usuário. Ao identificar que o item possui Acesso Restrito, nossa customização verifica se o usuário em questão possui diretrizes de acesso para o item, em caso negativo, o usuário recebe uma mensagem informando que não possui o acesso necessário e apenas os metadados são disponibilizados. Caso o usuário possua as credenciais, o fluxo do DSpace é desviado, para uma rotina no qual o arquivo encriptado é consultado, a chave de decriptação é usada para decodificar o arquivo e uma cópia legível é entregue temporariamente por 30 minutos ao usuário, posteriormente a cópia é deletada, e se o usuário precisar do item, todo o processo de decriptação será feito novamente.

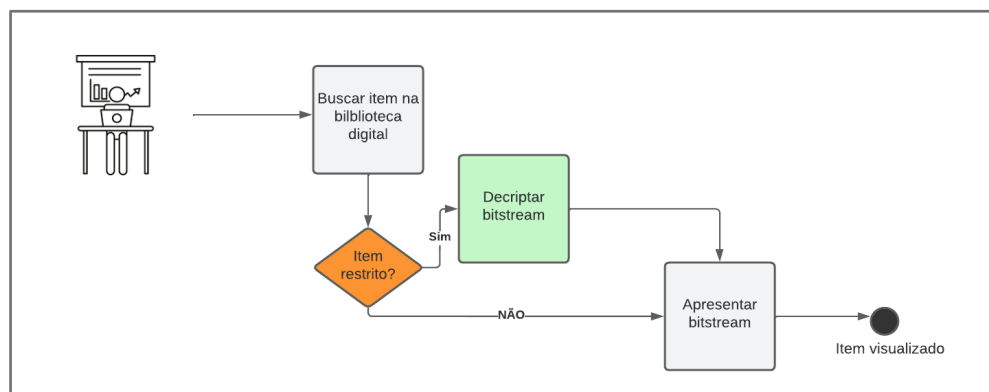


FIGURA 6. Fluxo de recuperação do item

Fonte: Elaboração dos autores, 2023

Uma observação, se faz necessário, a customização não usa chave de encriptação e decriptação para cada usuário, mas sim, em função de um grupo. Na customização implementada, é necessário um grupo para manter usuários com acesso a itens restritos na biblioteca digital. Através do serviço de controle de acesso em grupos disponibilizado pelo DSpace, um grupo chamado SAE foi criado para gerenciar os usuários que possuem acesso a itens restritos.

A configuração do grupo e da chave secreta é informada no `dspace.cfg/local.cfg` nas variáveis `dspace.group` e `dspace.key`, respectivamente. Os usuários pertencentes ao grupo, possuem livre arbítrio para consultar documentos confidenciais da biblioteca digital em questão. Vale ressaltar que, usuários pertencentes ao grupo administrador não necessariamente possuem acesso a itens restritos, para efetivar, é necessário a inclusão no grupo SAE. Além disso, acessar o host no qual o DSpace está implantado não dará ao invasor acesso a itens restritos, pois o fluxo de bits armazenado no diretório `assetstore` será criptografado e ilegível a olho nu.

#### 4.4 Testes

Os testes foram conduzidos em um ambiente com as seguintes características:

- Sistema operacional linux/ubuntu versão 20.04
- Processador quad-core
- 8GB de memória RAM
- 100 GB de espaço em disco

A fase de testes foi elaborada em 2 ambientes: desenvolvimento e produção. Informáticos e um bibliotecário foram envolvidos no processo de testes de validação. Quadro 01, expõe os testes realizados com os tipos de bitstreams mais comuns numa biblioteca digital. Estes testes foram realizados diretamente por meio das interfaces do DSpace com vários tipos de usuários, de forma a verificar a efetividade das implementações.

<b>TIPO BITSTREAM</b>	<b>PERFIL DE USUÁRIO</b>	<b>PERTENCE AO GRUPO SAE?</b>	<b>RESULTADO</b>
.pdf, .txt, .doc, .docx	usuários do grupo administrador	não	impedimento de acesso ao arquivo e aviso.
.pdf, .txt, .doc, .docx	usuários do grupo administrador	sim	Disponibilidade do arquivo decriptado.
.pdf, .txt, .doc, .docx	usuário anônimo	não	impedimento de acesso ao arquivo e aviso.
.pdf, .txt, .doc, .docx	usuários pertencentes a grupos diferentes do grupo administrador	sim	Disponibilidade do arquivo decriptado.
.pdf, .txt, .doc, .docx	usuários pertencentes a grupos diferentes do grupo administrador	não	impedimento de acesso ao arquivo e aviso.

QUADRO 1. Testes realizados

Fonte: Elaboração dos autores, 2023

A coluna "PERTENCE AO GRUPO" refere-se ao grupo que contém todos os usuários com permissão de acesso a depósitos restritos. É importante ressaltar que, mesmo que um usuário sem permissão acesse um item restrito, os metadados do item serão listados, juntamente com a informação de que o item é restrito. No entanto, o arquivo em si não estará disponível para download ou leitura, pois o DSpace possui leitores de arquivos textuais integrados.

Outro ponto importante a destacar é que a permissão de acesso a itens restritos não depende do tipo de perfil do usuário. Ser qualificado para acessar documentos restritos é o único requisito para obter acesso. No entanto, um usuário com perfil de administrador pode se incluir automaticamente no grupo de acesso a documentos restritos. No entanto, caso ele não execute essa etapa de customização, ele não será considerado apto para acessar documentos restritos.

O impacto da performance do serviço no DSpace não foi mensurado, todavia, a metodologia usada não apresenta fragmento de código que possa comprometer a ponto de comprometer a usabilidade. Além disso, enfatizamos que a biblioteca de encriptação que usamos é conceituada na literatura não abrindo brechas futuramente para acesso a hospedeiros maus intencionados.

Validações com relação aos serviços padrões do DSpace: busca simples e avançada; navegação por data, autor, assunto, título foram verificadas, e nenhuma inconsistência foi diagnosticada. O comportamento já era esperado, já que houve uma preocupação em desenvolver um serviço modular fracamente acoplado sem interferência com os demais serviços.

Em acesso via servidor, diretamente no *assetstore*, diretório em que os arquivos são carregados, os arquivos encriptados estavam ilegíveis. Com isso, foi possível verificar como o processo ocorreu de forma a atender a necessidade de codificar os documentos sensíveis, de forma não deixá-los em claro. Assim, caso haja algum acesso indevido, o conteúdo dos documentos estarão preservados do acesso.

Com relação aos demais módulos do software, por exemplo o protocolo OAI-ORE, a coleta de itens encriptados é possível, contudo, mantendo a integridade da solução proposta, esses arquivos não serão legíveis ao coletador, a menos que ele possua a chave para realizar a decifração.

Por fim, destaca-se que os testes realizados foram em arquivos de no máximo 100MB. Com relação a performance de desempenho, conforme discutido em Diffie e Hellman (2022) e Abood e Guirguis (2018), o AES é um dos melhores na literatura. Os testes realizados mostram que a encriptação e decríptação não impõe um overhead que seja notado pelo usuário, tornando o processo transparente para os usuários.

#### **4.5 Apresentação dos resultados**

Conforme foi apresentado, a implementação do processo de encriptação teve sucesso, possibilitando que os documentos indicados como restritos, ao serem carregados, fossem armazenados de forma codificada. Os documentos sem restrição são armazenados em claro, ou seja, legíveis a leitura, caso sejam acessados diretamente nos diretórios dos servidores.

Tanto no processo de depósito, quanto de recuperação, a encriptação/decriptação ocorreu de forma transparente ao usuário. No processo de depósito, o arquivo é encriptado de forma transparente ao usuário e, no caso de acessar o arquivo no servidor, no diretório assetstore, o arquivo estará indecifrável. No processo de recuperação, seja via busca ou pela navegação, o arquivo é decodificado e apresentado ao usuário na sua forma legível.

Assim, o processo de implementar a criptografia no DSpace se mostrou eficaz, podendo ser implementado, de forma a promover maior segurança contra ataques hackers.

#### **5 Considerações finais**

A segurança tem se apresentado como tônica nos sistemas de informação que utilizam ferramentas informatizadas e fazem gestão de documentos em formato digital. Tanto que, em alguns casos, esse ponto é estratégico, como forma de apresentar restrições ao acesso indevido. Em outros casos, a preocupação é com a integridade das informações armazenadas em sistemas

de informação. De qualquer forma, com o uso cada vez maior da internet por todos, com documentação disponível para acesso, a segurança torna-se essencial.

Nesse sentido, a segurança digital em organizações deve ser parte da governança, parte de estratégias de um plano de segurança, incluindo o letramento sobre segurança digital. Proteger o capital intelectual da organização, composta de documentação digital, é considerado como atividade vital contra-ataques oriundos do ambiente cada vez mais conectado.

Desta forma, implementar criptografia no DSpace atende, em parte, um plano de segurança, principalmente, para órgãos de governos que o utilizam para criar as suas bibliotecas digitais, principalmente por seu acervo ser constituído quase que totalmente de primeira fonte. Mesmo para repositórios de institutos de ensino e pesquisa, parte do acervo, como as teses e dissertações, são de primeira fonte, requerendo proteção contra-ataques.

## Bibliografia

- ABOOD, O. G., & GUIRGUIS, S. K. (2018). A Survey on Cryptography Algorithms. *International Journal of Scientific and Research Publications (IJSRP)*, 8(7). DOI: <[doi:10.29322/IJSRP.8.7.2018.p7978](https://doi.org/10.29322/IJSRP.8.7.2018.p7978)>
- ANGELIS, G. DE, FERRARI, A., GNESI, S., & POLINI, A. (2018). Requirements elicitation and refinement in collaborative research projects. *Journal of Software: Evolution and Process*, 30(12), e1990. DOI: <[doi:10.1002/smr.1990](https://doi.org/10.1002/smr.1990)>
- AUFA, F. J., ENDROYONO, & AFFANDI, A. (2018). Security System Analysis in Combination Method: RSA Encryption and Digital Signature Algorithm. *2018 4th International Conference on Science and Technology (ICST)*, 1–5. Yogyakarta: IEEE. DOI: <[doi:10.1109/ICSTC.2018.8528584](https://doi.org/10.1109/ICSTC.2018.8528584)>
- BASEVI, T. (2005). BDJur Consortium: Juridical Digital Library: Implementing DSpace in the Brazilian Judiciary. *Proceedings*, 127–132. Leuven-Heverlee, Bélgica: Peeters Publishing Leuven. <<https://elpub.architexturez.net/doc/oai-elpub-id-150elpub2005>>
- BJÖRK, B.-C. (2007). A model of scientific communication as a global



- distributed information system. *Information Research: an International Electronic Journal*, 12(2). <<http://informationr.net/ir/12-2/paper307.html>>
- DIFFIE, W., & HELLMAN, M. E. (2022). New Directions in Cryptography. Em R. Slayton (Org.), *Democratizing Cryptography* (1º ed, p. 365–390). New York, NY, USA: ACM. DOI: <[doi:10.1145/3549993.3550007](https://doi.org/10.1145/3549993.3550007)>
- HERON, S. (2009). Advanced Encryption Standard (AES). *Network Security*, 2009(12), 8-12. DOI: <[doi:10.1016/S1353-4858\(10\)70006-4](https://doi.org/10.1016/S1353-4858(10)70006-4)>
- MACÊDO, D. J., SHINTAKU, M., & BRITO, R. F. DE. (2015). Dublin Core usage for describing documents in Brazilian Government Digital Libraries. *Anais*, 129-135. São Paulo.
- MENEZES, A., & STEBILA, D. (2021). The Advanced Encryption Standard: 20 Years Later. *IEEE Security & Privacy*, 19(6), 98–102. DOI: <[doi:10.1109/MSEC.2021.3107078](https://doi.org/10.1109/MSEC.2021.3107078)>
- PAAR, C., & PELZL, J. (2010). Introduction to Cryptography and Data Security. Em C. Paar & J. Pelzl, *Understanding Cryptography* (p. 1–27). Berlin, Heidelberg: Springer Berlin Heidelberg. DOI: <[doi:10.1007/978-3-642-04101-3\\_1](https://doi.org/10.1007/978-3-642-04101-3_1)>
- SELENT, D. (2010). Advanced Encryption Standard. *InSight: Rivier Academic Journal*, 6(2). <<https://www2.rivier.edu/journal/roaj-fall-2010/j455-selent-aes.pdf>>
- SHANNON, C. E. (1949). Communication theory of secrecy systems. *Bell System Technical Journal*, 28(4), 656-715. DOI: <[doi:10.1002/j.1538-7305.1949.tb00928.x](https://doi.org/10.1002/j.1538-7305.1949.tb00928.x)>
- SIMMONS, G. J. (1979). Symmetric and Asymmetric Encryption. *ACM Computing Surveys*, 11(4), 305-330. DOI: <[doi:10.1145/356789.356793](https://doi.org/10.1145/356789.356793)>
- WAZLAWICK, R. S. (2014). *Metodologia de pesquisa para ciência da computação* (2º ed). Rio de Janeiro: Elsevier. <<http://www.sciencedirect.com/science/book/9788535277821>>
- YASEEN, M., & FAROOQ, U. (2018). Requirement Elicitation Model (REM) in the Context of Global Software Development. *Global Journal of Computer Science and Technology*, 18(C2), 37-41.
- YASSEIN, M. B., ALJAWARNEH, S., QAWASMEH, E., MARDINI, W., & KHAMAYSEH, Y. (2017). Comprehensive study of symmetric key and asymmetric key encryption algorithms. *2017 International Conference on Engineering and Technology (ICET)*, 1-7. Antalya: IEEE. <[doi:10.1109/ICEngTechnol.2017.8308215](https://doi.org/10.1109/ICEngTechnol.2017.8308215)>

# Analysis of the Open Data Landscape in Mexico

## **JOANNA ALVARADO URIBE**

INSTITUTE FOR THE FUTURE OF EDUCATION

TECNOLOGICO DE MONTERREY

[joanna.alvarado@tec.mx](mailto:joanna.alvarado@tec.mx)

## **PAOLA MEJÍA ALMADA**

INSTITUTE FOR THE FUTURE OF EDUCATION

TECNOLOGICO DE MONTERREY

[gabriela.almada@tec.mx](mailto:gabriela.almada@tec.mx)

## **ALMA BEATRIZ RIVERA AGUILERA**

BIBLIOTECA FRANCISCO XAVIER CLAVIGERO

UNIVERSIDAD IBEROAMERICANA

[alma.rivera@ibero.mx](mailto:alma.rivera@ibero.mx)

## **BENJAMÍN ALEJANDRO GUERRERO OLVERA**

BIBLIOTECA FRANCISCO XAVIER CLAVIGERO

UNIVERSIDAD IBEROAMERICANA

[benjamin.olvera@ibero.mx](mailto:benjamin.olvera@ibero.mx)

## **MARÍA TERESA VILLALÓN GUZMÁN**

INSTITUTO TECNOLÓGICO DE CELAYA

TECNOLÓGICO NACIONAL DE MÉXICO

[teresa.villalon@itcelaya.edu.mx](mailto:teresa.villalon@itcelaya.edu.mx)

## **MARÍA GUADALUPE VEGA DÍAZ**

BIBLIOTECA DANIEL COSÍO VILLEGAS

EL COLEGIO DE MÉXICO

[guvega@colmex.mx](mailto:guvega@colmex.mx)

## **JOAQUÍN GIMÉNEZ HÉAU**

DIRECCIÓN GENERAL DE REPOSITARIOS UNIVERSITARIOS

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

[joaquin@dgru.unam.mx](mailto:joaquin@dgru.unam.mx)

### **EJE TEMÁTICO**

Datos abiertos

### **ABSTRACT**

The concept of open data originated from the idea of making government data available to anyone. In 2013, the Mexican government launched the National Digital Strategy to encourage the adoption and development of information by implementing five key enablers, including an open data portal to make government data more accessible to citizens. In 2014, the beta version of the [datos.gob.mx](https://datos.gob.mx) portal was launched. Due to the significant potential of data to generate benefits, it has been referred to as the new gold. However, open data also presents some challenges, such as lack of interoperability and reproducibility and the resources needed to provide and maintain open data available to all. Therefore, based on the transversal agenda REMERI-ANUIES-EduTraDi focused on Open Science Ecosystems, the objectives of this research are to identify and analyze institutional repositories promoting Open Data in Mexico in order to document their characteristics and determine possible areas for improvement in such initiatives. This investigation allowed the identification of ten repositories promoting open data in Mexico and the integration of 16 characteristics that helped to describe and analyze the open data landscape, showing that the different open data management efforts carried out by Mexican institutions are mostly consolidated.

### **KEYWORDS**

Open data; repository; Open Data Landscape; Mexico.

## Introduction

The concept of open data originated from the idea of making government data available to anyone, as government institutions collect a vast amount of information from a variety of sources. Such a concept gained momentum in the late 2000s due to the rise of the Internet and the proliferation of digital information. During this period, there was a concerted effort by governments and various organizations to provide access to these resources to a large number of users (The World Bank Group, Open Data in 60 Seconds, 2021). The first government policies appeared in 2009, while in 2010, the World Bank organization launched the first “Open Data Initiative”. This initiative had the goal of finding solutions to complicated development problems by giving researchers access to a wide range of data related to the global economy (The World Bank Group, Learning About the Open Data Initiative, 2012).

In addition to the Open Data Initiatives launched every year, more than 250 countries along with almost 50 developed and developing countries, as well as other organizations (such as the United Nations), have introduced Open Data initiatives (The World Bank Group, Open Data in 60 Seconds, 2021). In 2013, the Mexican government launched the National Digital Strategy to encourage the adoption and development of information by implementing five key enablers, including an open data portal to make government data more accessible to citizens (Digital Government Unit, 2015). This became a reality in 2014 when the beta version of the datos.gob.mx portal was launched. Since then, this portal has been available and has gone through several updates (Coordinación de Estrategia Digital Nacional, 2015).

The definition of open concerning open data and content states that: Open means that anyone can freely access, use, modify, and share it for any purpose (Open Knowledge Foundation, 2015). Moreover, it is also important to emphasize that open data must be available in a common, machine-readable format and licensed so that anyone can manipulate data. For example, transforming, combining, and sharing it with others, even commercially

(European Data, n.d.). Due to the significant potential of data to generate benefits, it has been referred to as the new gold. Some of the benefits of providing and using open data are transparency, public service improvement, innovation and economic value, and efficiency (The World Bank Group, Open Data in 60 Seconds, 2021). However, open data also presents some challenges, such as loss of data privacy and security, lack of interoperability and reproducibility, improvement of data quality, intellectual property rights, and the resources needed to provide and maintain open data available to all (Goben & Sandusky, 2020).

One of the most recent studies related to open data repositories highlights the importance of the role of infrastructure, as it becomes crucial in the dissemination, visibility, and openness of data for research processes (Science, et al., 2022). Therefore, based on the transversal agenda REMERI-ANUIES-EduTraDi focused on Open Science Ecosystems, as well as on the aforementioned background and related work, the objectives of this research are to identify and analyze institutional repositories promoting Open Data belonging to the public and private sectors in Mexico in order to document their characteristics and determine possible areas for improvement in such initiatives.

The rest of the article is organized as follows. Section 2 presents the methodology followed in this research project. Subsequently, Section 3 provides the results, analysis, and discussions about them. Finally, Section 4 gives the conclusions and proposals for improvement.

## **Methodology**

The identification and analysis of institutional repositories that incorporate Open Data belonging to the public and private sectors at the national level are intended to document their characteristics and determine areas for

improvement in such initiatives. To achieve this purpose, the methodology presented in Figure 1 and described below was followed:

- 1) Initial identification of repositories incorporating open data in Mexico. This first phase was based on two aspects: 1) on the knowledge and experience of the members of the research project, whose profiles are mainly related to the management of data repositories in Mexican institutions, and 2) on the search for repositories whose description or keywords included references to open data through different digital resources, such as websites, web platforms, research and dissemination publications, reports, among others.
- 2) Definition of the characteristics (variables) reviewed and registered for each repository identified in this investigation. To carry out this definition, the characteristics were established considering the purpose of the research, as well as different standards and guides related to types of data and areas of knowledge.
- 3) Collection of values for each characteristic of the repositories integrating open data in Mexico. To document the defined characteristics, the information provided on the platforms or websites of each repository was consulted and the values obtained were registered in a spreadsheet structured for this purpose. In addition, in some cases where the information was not available, the person responsible for the repository was contacted to request their support in providing the missing information.
- 4) Discussion of the information collected for each repository by the person responsible for its registration and all the members of the research project. This activity allowed to resolve doubts about the registered information and thus improve the quality and standardization of the information collected.
- 5) Elaboration of a quantitative analysis of the characteristics of the repositories. In order to discover the most used aspects and the areas for

improvement in the repositories promoting open data, statistics of the values and different visualizations by characteristic were carried out. Subsequently, the analysis of these representations and the information itself was performed.

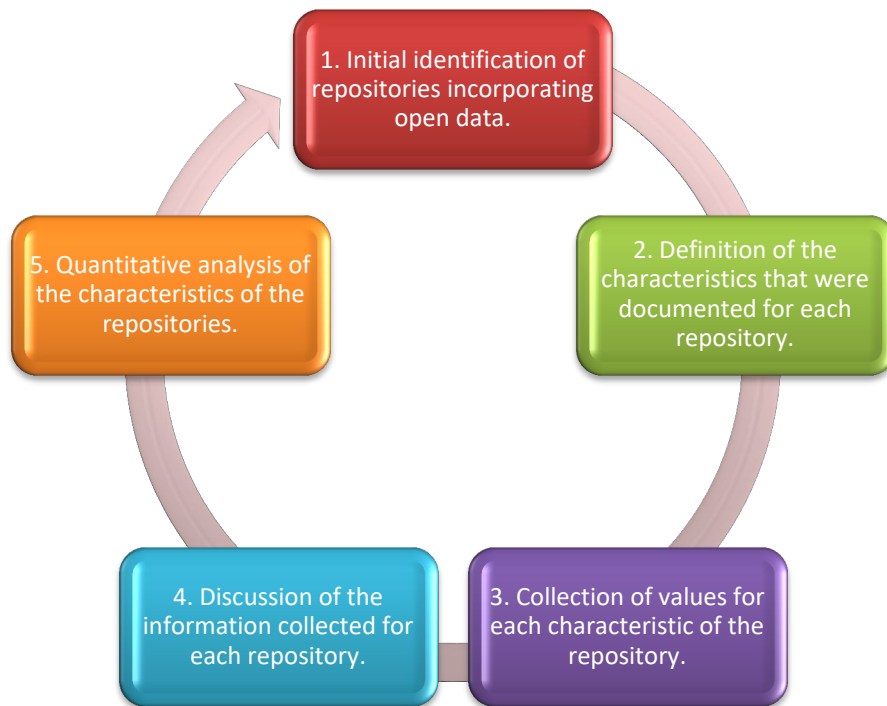


FIGURE 1. Methodology used to analyze the open data landscape in Mexico. Own elaboration.

## Results

Firstly, the characteristics considered in the documentation of the repositories promoting open data, partially or completely, in Mexico are described. Later, the information about the identified repositories and the graphs visualizing the collected characteristics are shown. Finally, the analysis of the information presented is provided.

### **Definition of the Collected Characteristics**

The 16 characteristics defined for the analysis of the repositories, promoting open data in Mexico, are related to their identification (such as name and Uniform Resource Locator (URL)), the area of knowledge of their content, their typology, numerals, metadata standards, use licenses, among other aspects oriented to data and software. The detailed list of features and their description is given below:

- 1)** Repository name.
- 2)** Repository URL.
- 3)** Description of the repository. Overview of the repository content.
- 4)** Institution. Institution name responsible for the repository.
- 5)** Sector. The sector to which the institution responsible for the repository belongs is indicated.
- 6)** Name of the person in charge. Full name of the person responsible for the repository or of the person related to the activities of the repository.
- 7)** Email of the person in charge.
- 8)** Data typology (LEARN project, 2020).
  - a)** By its type. It indicates whether the data is in the type of electronic text documents, data sheets, metadata, among others.
  - b)** By its file formats: It is specified if the data is textual, numerical, structured, among others, as well as the file extensions, for example, .csv, .txt, .pdf.
  - c)** By its data processing level: the data can be indicated as 1) raw or unprocessed, 2) processed, or 3) analyzed.
  - d)** By the data generation source: the data can be considered as 1) canonical or reference, 2) experimental, 3) models or simulations, 4) derived or compiled, or 5) Observational.
- 9)** Classification based on the data structure. It indicates whether the repository is structured into Collections, Datasets, Geospatials, or Other.



- 10) Areas of Knowledge. The area of knowledge of the repository content is indicated according to the areas defined by the Consejo Nacional de Ciencia y Tecnología (CONACYT) (CONACYT, 2022).
- 11) Metadata standard.
- 12) Technology/Platform. The platform hosting the repository is specified.
- 13) Data use license.
- 14) Data use restriction.
- 15) Numerals. The figure (obtained or calculated) of collections and resources in the repository is provided.
- 16) Open Data. It indicates whether the repository contains only open data.

### ***Repositories Promoting Open Data in Mexico***

The identification of repositories promoting open data at the national level was carried out from September 9, 2022, to November 2, 2022. To consider the integration of the repository into this research, the following aspects were taken into account:

- 1) Repositories must present at least one open data record.
- 2) Open data must come from Mexican institutions.
- 3) In case of that an institutional repository indicates that its open data is in an exclusive open data repository, the repository that should be considered for the analysis is the one that only integrates open data.

Based on these considerations, a total of ten repositories presenting open data were selected to carry out their complete documentation. The overview of these repositories can be seen in Table 1, which is updated to the review period of the characteristics from May 8 to 12, 2023.

No	REPOSITORY NAME	REPOSITORY URL	INSTITUTION	SECTOR	OPEN DATA EXCLUSIVE
1	Portal de Datos Abiertos UNAM	<a href="https://datosabiertos.unam.mx/">https://datosabiertos.unam.mx/</a>	Universidad Nacional Autónoma de México (UNAM) / Dirección General de Repositorios Universitarios	Public	Yes
2	Data Hub del Tecnológico de Monterrey	<a href="https://datahub.tecnologico.mx/">https://datahub.tecnologico.mx/</a>	Tecnológico de Monterrey	Private	Yes
3	Datos Abiertos de México	<a href="https://datos.gob.mx/">https://datos.gob.mx/</a>	Gobierno de México	Public	Yes
4	Laboratorio Nacional de Políticas Públicas	<a href="http://datos.cide.edu/community-list">http://datos.cide.edu/community-list</a>	Laboratorio Nacional de Políticas Públicas	Public	Yes
5	CIMMYT Research Data & Software Repository Network	<a href="https://data.cimmyt.org/dataverse/root?q=&amp;types=dataverses&amp;sort=dateSort&amp;order=asc&amp;page=1">https://data.cimmyt.org/dataverse/root?q=&amp;types=dataverses&amp;sort=dateSort&amp;order=asc&amp;page=1</a>	International Maize and Wheat Improvement Center	Public and private	Yes
6	Repositorio Digital del Servicio Sismológico Nacional	<a href="http://www2.ssn.unam.mx:8080/catalogo/">http://www2.ssn.unam.mx:8080/catalogo/</a>	UNAM / Servicio Sismológico Nacional / Instituto de Geología	Public	Yes
7	Datos abiertos - INEGI	<a href="https://www.inegi.org.mx/datosabiertos/">https://www.inegi.org.mx/datosabiertos/</a>	Instituto Nacional de Estadística y Geografía (INEGI)	Public	Yes
8	Repositorio de Documentación sobre Desapariciones en México (RDDM)	<a href="https://rddm.mx/">https://rddm.mx/</a>	Center for Research Libraries; El Colegio de México; Universidad Iberoamericana; UNAM / Instituto de Investigaciones	Public	Partially

<b>NO</b>	<b>REPOSITORY NAME</b>	<b>REPOSITORY URL</b>	<b>INSTITUTION</b>	<b>SECTOR</b>	<b>OPEN DATA EXCLUSIVE</b>
			Jurídicas		
9	Datos Abiertos - SSA/DGIS	<a href="http://www.dgis.salud.gob.mx/contenidos/basesdedatos/Datos_Abiertos_gobmx.html">http://www.dgis.salud.gob.mx/contenidos/basesdedatos/Datos_Abiertos_gobmx.html</a>	Secretaría de Salud (SSA) / Dirección General de Información en Salud (DGIS)	Public	Yes
10	Repositorio IBERO	<a href="https://ri.iberomex.mx/">https://ri.iberomex.mx/</a>	Universidad Iberoamericana	Private	Partially

TABLE 1. Description of the ten repositories promoting open data in Mexico.

### ***Open Data Landscape in Mexico***

As can be seen in Table 1, ten repositories promoting open data in Mexico were identified. Of these, more than half correspond to public institutions (such as universities and government institutions) and two to private educational institutions. In addition, 80 % of the identified repositories exclusively publish open data mainly in the areas of knowledge of "Social Sciences", "Physical-Mathematical and Earth Sciences", and "Engineering and Technological Development". Considering the number of collections and data records (resources), some repositories can be expected to have up to 40 collections and 2,106,655 data records, while some have at least one collection and six resources.

Regarding the data typology, Figure 2 shows that the data is mainly provided in Datasheets (22.73 %), followed by Electronic text documents (20.45 %), and also with their Metadata (13.64 %). This has a relationship to Figure 3, where the .csv, .pdf, .xlsx, and .txt file formats are more represented (larger) in the word cloud and therefore, more published in open data repositories.

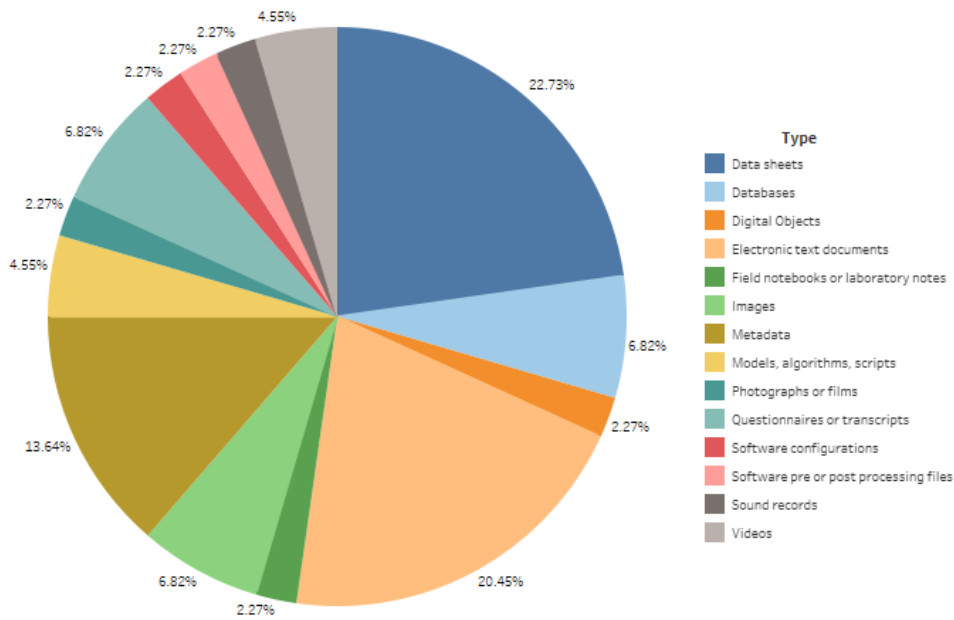


FIGURE 1. Data typology according to its type. Own elaboration



FIGURE 2. Data typology according to its file formats. Own elaboration

From Figures 4 and 5, it can be seen that all repositories contain processed data, followed by analyzed data (90 %), generated mainly from other data (derived or compiled) and considered as curated datasets (canonical or reference).

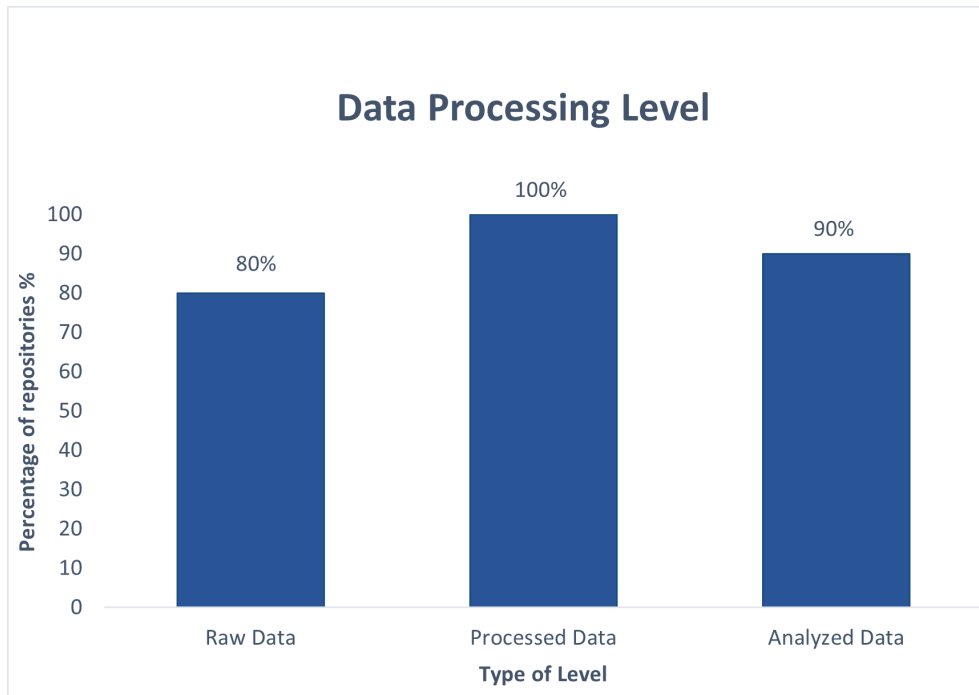


FIGURE 3. Data typology according to its data processing level. Own elaboration

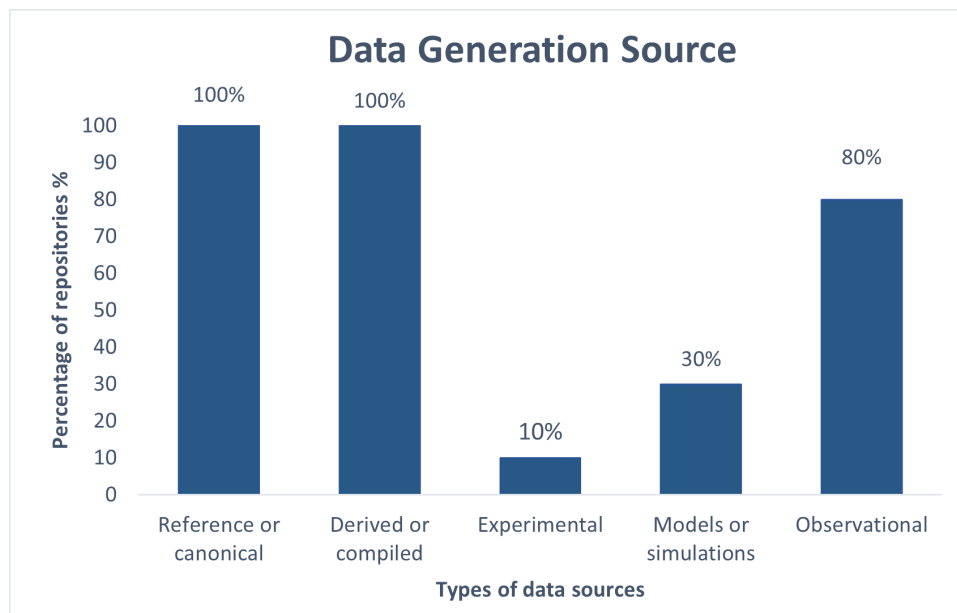


FIGURE 4. Data typology according to its data generation source. Own elaboration

On the other hand, the Dublin Core Metadata Initiative (DCMI) is the most followed by repositories by incorporating its Dublin Core (DC) vocabulary. However, in terms of the technology or platform hosting the repositories, there

is no adoption trend in the ten repositories reviewed since Dataverse and DSpace present the same percentage of use (20 %). In addition to the fact that there are repositories where the information on their platform or technology is not provided (20 %), as can be seen in Figure 6.

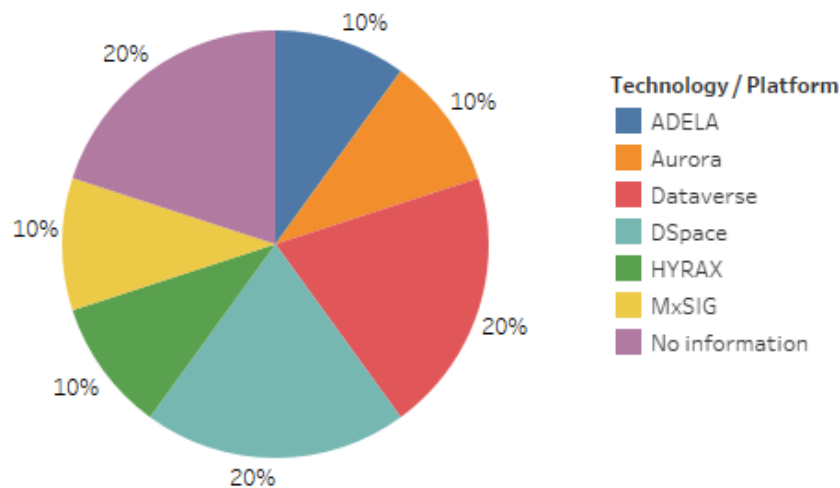


FIGURE 5. The platform or technology hosting the repository. Own elaboration

Finally, the widely popular Creative Commons (CC) licenses were the most widely used in repositories. However, some restrictions of use and access were found, such as 1) to freely use the data, the source of origin must be cited, 2) limitation of access to resources (which can be accessed upon request), and 3) data embargo.

## Conclusions

This investigation allowed the identification of ten repositories promoting open data in Mexico and the integration of 16 characteristics that helped to describe and analyze the landscape of open data published in said repositories until May 12, 2023. The results of this research show that the different open data management efforts carried out by Mexican institutions are mostly consolidated since they incorporate metadata standards, platforms, and use

licenses that enable free data access, use, and sharing. However, some areas for improvement were identified, which are organized into four proposals:

- 1)** Establish an agreement on the information to be provided and its structuring for a better description, analysis, and evaluation of the data and the repositories. Some recommendations are:
  - a)** Enrich the documentation to establish record guidelines and examples of the different data types. For example, in geospatial data, it is suggested to add the coordinates.
  - b)** Provide a base of descriptive characteristics of the repository. For example, the metadata standards followed and the platform or technology hosting the repository.
  - c)** Define the same classification of resources by areas of knowledge, type, file formats, among other aspects, such as those mentioned in this investigation, for their correct review and analysis. These categories could be presented as content filters.
- 2)** Promote the integration of open data in a unique repository or in different repositories where it is sought to unify metadata standards and the hosting technology or platform to allow:
  - a)** Interoperability among data.
  - b)** Informed decision-making regarding an area or topic of interest based on the available data.
- 3)** Strengthen the adoption of laws and agreements on the protection of personal data and the treatment of sensitive data. In this way, although it seeks to promote and facilitate access to data and its documentation, possible inappropriate use of them is being prevented, especially when the integrity of people may be affected.
- 4)** Encourage the integration of more private institutions in the Open Data initiative of Mexico in order to promote the publication of their data and, in this way, support research boosted in favor of improving the quality of life in the country and the world.

As future work, it is proposed to continue completing the information that could not be obtained from the repository to ensure that all the characteristics defined in this investigation are documented. In addition, carry out an evaluation of these repositories through an evaluation standard widely used and reported in the literature.

ACKNOWLEDGMENTS. The authors thank Gerardo Castañeda Garza for his collaboration in the early phases of the project, as well as Rosalina Vázquez Tapia, Antonio Razo, and Martín Adalberto Tena Espinoza de los Monteros for the coordination and support within the framework of the transversal agenda REMERI-ANUIES-EduTraDi to which this research belongs.

## References

- CONACYT, C. (2022). Términos de referencia. Convocatoria de Ciencia Básica y/o Ciencia de Frontera: <[https://conacyt.mx/wp-content/uploads/convocatorias/ciencia\\_de\\_frontera/paradigmas\\_y\\_controversias/2022/TdR\\_Paradigmas\\_y\\_Controversias\\_de\\_la\\_Ciencia\\_2022\\_VF.pdf](https://conacyt.mx/wp-content/uploads/convocatorias/ciencia_de_frontera/paradigmas_y_controversias/2022/TdR_Paradigmas_y_Controversias_de_la_Ciencia_2022_VF.pdf)>
- COORDINACIÓN DE ESTRATEGIA DIGITAL NACIONAL. (2015). Datos.gob.mx: Impulsa los Datos Abiertos en Mexico. <<https://datos.gob.mx/blog/datosgobmx-impulsa-los-datos-abiertos-en-mexico?category=noticias&tag=desarrollo>>
- DIGITAL GOVERNMENT UNIT, M. O. (2015). Digital government toolkit. National Digital Strategy. <<https://www.oecd.org/gov/mexico-digital-strategy.pdf>>
- EUROPEAN DATA. (n. d.). What is open data? <<https://data.europa.eu/elearning/en/module1/#/id/co-01>>
- GOBEN, A., & SANDUSKY, R. (2020). Open data repositories. Current risks and opportunities. <<https://crln.acrl.org/index.php/crlnews/article/view/24273/32092and#:~:text=Another%20challenge%20is%20the%20potential,continue%20to%20pursue%20grant%20funding>>
- LEARN PROJECT, L. A. (2020). Gestión de datos de investigación. <<https://biblioguias.cepal.org/gestion-de-datos-de-investigacion>>
- OPEN KNOWLEDGE FOUNDATION. (2015). Open Definition. <<https://opendefinition.org/>>



SCIENCE, D., GOODEY, G., HAHNEL, M., ZHOU, Y., JIANG, L., CHANDRAMOULISWARAN, I., DAY, L. (2022). The State of Open Data 2022. doi: <<https://doi.org/10.6084/m9.figshare.21276984.v5>>

THE WORLD BANK GROUP. (2012). Learning About the Open Data Initiative. <<https://www.worldbank.org/en/news/feature/2012/03/22/learning-about-the-open-data-initiative>>

THE WORLD BANK GROUP. (2021). Open Data in 60 Seconds. <<http://opendatatoolkit.worldbank.org/en/open-data-in-60-seconds.html>>

# Análisis de escenarios y protocolos para integración de repositorios digitales y sitios web institucionales

## **PABLO DE ALBUQUERQUE**

CESGI COMISIÓN DE INVESTIGACIONES CIENTÍFICAS (CIC)

[pablo@sedici.unlp.edu.ar](mailto:pablo@sedici.unlp.edu.ar)

## **PABLO GABRIEL TERRONE**

PREBI-SEDICI UNIVERSIDAD NACIONAL DE LA PLATA (UNLP)

[pabloterrone@sedici.unlp.edu.ar](mailto:pabloterrone@sedici.unlp.edu.ar)

## **GONZALO LUJÁN VILLARREAL**

PREBI-SEDICI UNIVERSIDAD NACIONAL DE LA PLATA (UNLP)

CESGI COMISIÓN DE INVESTIGACIONES CIENTÍFICAS (CIC)

[gonzalo@prebi.unlp.edu.ar](mailto:gonzalo@prebi.unlp.edu.ar)

## **MARISA RAQUEL DE GIUSTI**

PREBI-SEDICI UNIVERSIDAD NACIONAL DE LA PLATA (UNLP)

CESGI COMISIÓN DE INVESTIGACIONES CIENTÍFICAS (CIC)

[marisa.degiusti@sedici.unlp.edu.ar](mailto:marisa.degiusti@sedici.unlp.edu.ar)

## **EJE TEMÁTICO**

Infraestructura tecnológica

## **RESUMEN**

En este trabajo se analizan los protocolos de interoperabilidad OAI-PMH y OpenSearch, junto a la API REST disponible en DSpace 7, desde el punto de vista de su utilidad a la hora de exponer recursos de repositorios digitales en sitios web institucionales. Se analiza el grado de adopción de cada protocolo, su facilidad de integración con CMS como Wordpress o Drupal, y el nivel de personalización soportado en las consultas y en las respuestas

que generan. A partir de este análisis, se evalúan distintos escenarios en los cuales es posible utilizar alguno de los protocolos, y se ejemplifica con URL y criterios de filtrado en los casos donde sea posible. La evaluación de estos escenarios sugiere que, si bien OAI-PMH resulta adecuado para cosechas masivas, no parece una buena solución cuando se requieren pocos recursos, integrados en CMS, y con criterios de filtrado muy específicos. El protocolo OpenSearch permite una fácil integración con CMS, pero resulta muy limitado por la información que ofrece y la falta de estandarización. Finalmente, la API REST de DSpace 7 ofrece una buena alternativa a OpenSearch, pero su implementación es muy reciente y su integración con CMS más compleja.

#### **PALABRAS CLAVE**

Repositorios digitales; API REST; OAI PMH; OpenSearch; interoperabilidad.

### **Introducción**

Los repositorios digitales brindan numerosos mecanismos de interoperabilidad, y con ello buscan fomentar que los recursos alojados por estas plataformas sean integrados dentro de servicios y sistemas de terceros (De Giusti *et al.*, 2013). Esta integración puede realizarse con múltiples fines, como exponer recursos externos para un portal de tesis, un agregador de recursos regional o un blog personal, o como fuente de datos para poblar o agilizar la carga en otro repositorio digital o un CRIS. También sirven para brindar servicios de análisis de producción científica y/o integración de métricas, o incluso para proveer de recursos a sistemas de minería de textos o aprendizaje automático. El grado de detalle con el que se catalogan los recursos en los repositorios digitales, sumado a los estándares de normalización de metadatos utilizados, generan un entorno ideal para que muchos sistemas los aprovechen como fuentes primarias de información.

Las instituciones académicas y científicas cuentan con sitios web institucionales en los cuales describen, con mayor o menor grado de detalle, su organización y estructura interna, los objetivos que persiguen y el alcance de sus funciones. Muchas instituciones también utilizan sus sitios web como

espacios para la difusión de sus actividades, como noticias, resoluciones, agenda de eventos y logros o premios, y del trabajo que realizan sus miembros, por lo general a través de publicaciones científicas, presentaciones realizadas en congresos, tesis defendidas, patentes registradas, libros publicados e informes de proyectos de investigación. El mantenimiento y actualización periódica del contenido de los sitios web requiere contar con personal capacitado en comunicación institucional y científica; asimismo, a fin de asegurar una cobertura informativa que dé cuenta del trabajo realizado por todas las áreas de la institución, se requiere un compromiso de todos los espacios para comunicar sus actividades y su progreso al área responsable de la gestión del sitio web.

La integración de los repositorios digitales con sitios web institucionales abre las puertas a las instituciones académicas y científicas para ampliar y optimizar la forma de mostrar al mundo la producción de sus propios investigadores. Una mayor exposición podría incidir en el impacto de dicha producción tanto en la comunidad científica como en la sociedad en general. Esta integración puede aplicarse en diferentes contextos, que estarán sujetos a los objetivos que cada institución persigue con su repositorio institucional: tipo de recursos que aloja, estructura de colecciones y comunidades definidas, protocolos de interoperabilidad disponibles, formatos de representación de los recursos, capacidades de consulta y filtrado de datos, entre otros.

### **Protocolos de interoperabilidad en repositorios digitales**

Para poder integrar contenidos de los repositorios digitales con los sitios web institucionales, es necesario contar con mecanismos que permitan recuperar recursos desde los repositorios, y mostrarlos en los sitios web institucionales. Dichos sitios podrían estar basados en desarrollos informáticos a medida, aunque por lo general son implementados a través de Sistemas de Gestión de Contenidos (CMS) como Wordpress, Joomla! o Drupal. Esta no es

una característica particular de las instituciones académicas, ya que buena parte de los sitios web del mundo se encuentran desarrollados sobre CMS (2023's CMS Market Share Report – Latest Trends and Usage Stats, 2023; Open-source CMS Market Share, 2023), entre los que se destaca Wordpress como el más utilizado a nivel mundial, muy por encima del resto, y los sistemas Drupal y Joomla! dentro de los 10 más utilizados (Tabla 1 y Gráfico 1)

	2022 1 May	2022 1 Jun	2022 1 Jul	2022 1 Aug	2022 1 Sep	2022 1 Oct	2022 1 Nov	2022 1 Dec	2023 1 Jan	2023 1 Feb	2023 1 Mar	2023 1 Apr	2023 1 May	2023 8 May
WordPress	64.2%	64.1%	64.2%	64.3%	64.2%	64.3%	64.2%	64.0%	63.7%	63.5%	63.4%	63.3%	63.3%	63.3%
Shopify	6.5%	6.4%	6.3%	6.2%	6.2%	6.1%	6.1%	6.0%	5.7%	5.5%	5.5%	5.5%	5.6%	5.6%
Wix	3.4%	3.4%	3.4%	3.4%	3.4%	3.4%	3.4%	3.5%	3.6%	3.7%	3.7%	3.7%	3.7%	3.7%
Squarespace	3.0%	3.0%	3.0%	3.0%	3.0%	2.9%	2.9%	2.9%	3.0%	3.0%	3.1%	3.1%	3.1%	3.1%
Joomla	2.5%	2.5%	2.5%	2.5%	2.5%	2.5%	2.4%	2.5%	2.7%	2.7%	2.7%	2.7%	2.7%	2.7%
Drupal	1.9%	1.9%	1.8%	1.8%	1.8%	1.8%	1.8%	1.8%	1.8%	1.8%	1.8%	1.8%	1.8%	1.8%
Adobe Systems	1.6%	1.6%	1.6%	1.6%	1.6%	1.6%	1.6%	1.6%	1.7%	1.7%	1.7%	1.7%	1.6%	1.6%
Google Systems	1.4%	1.4%	1.4%	1.4%	1.3%	1.3%	1.3%	1.3%	1.2%	1.2%	1.2%	1.1%	1.1%	1.1%
PrestaShop	0.8%	0.8%	0.8%	0.8%	0.8%	0.8%	0.9%	0.9%	1.0%	1.1%	1.1%	1.1%	1.1%	1.1%
Bitrix	1.2%	1.2%	1.2%	1.2%	1.2%	1.1%	1.1%	1.1%	1.1%	1.1%	1.1%	1.1%	1.1%	1.1%
Webflow	0.8%	0.9%	0.9%	0.9%	0.9%	0.9%	0.9%	0.9%	0.9%	0.9%	0.9%	0.9%	0.9%	0.9%
OpenCart	0.8%	0.8%	0.8%	0.8%	0.8%	0.8%	0.8%	0.8%	0.8%	0.8%	0.8%	0.8%	0.8%	0.8%

TABLA 1. Tendencias de cuota de mercado para los sistemas de gestión de contenido.

Fuente: <[https://w3techs.com/technologies/history\\_overview/content\\_management](https://w3techs.com/technologies/history_overview/content_management)>

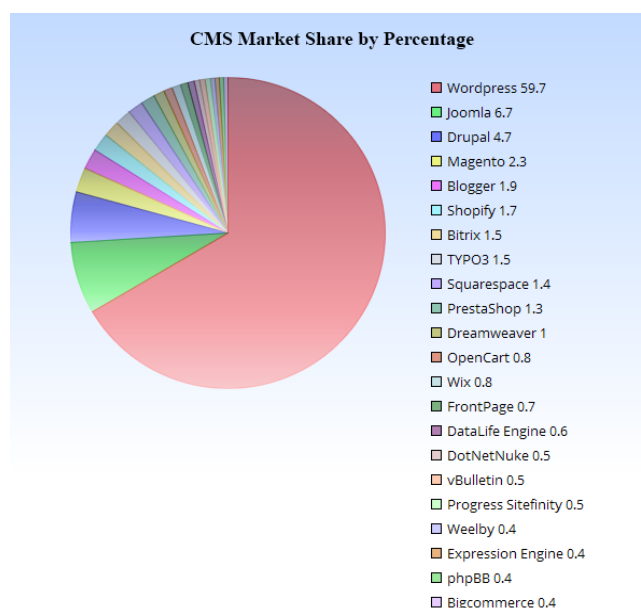


GRÁFICO 1. Resumen de todos los CMS con más del 0,4 % de cuota de mercado

Fuente <<https://www.opensourcecms.com/cms-market-share/>>

En esta sección se expondrán los protocolos de interoperabilidad más utilizados en el contexto de repositorios digitales y se los analizará como herramientas para exponer recursos en sistemas de terceros. En dicho análisis se definirán criterios como:

- forma de recuperación de información
- formato de intercambio de datos
- grado de estandarización del protocolo
- frecuencia de actualización de los datos involucrados

### **OAI-PMH: Open Archives Initiative - Protocol for Metadata Harvesting**

OAI-PMH es un protocolo estándar que permite la recolección y recuperación automatizada de metadatos de recursos digitales desde repositorios. El formato de respuesta utilizado por OAI-PMH es XML, generalmente bajo un esquema de metadatos estandarizado (Dublin Core o Marc, por ejemplo), lo que permite una fácil integración con otros sistemas y herramientas.

En cuanto al grado de estandarización, OAI-PMH es un estándar abierto y ampliamente utilizado en el ámbito de la educación y la investigación. Esto significa que los repositorios que implementan este protocolo deben cumplir con ciertas especificaciones y requisitos técnicos, y esto garantiza la interoperabilidad y la compatibilidad con otros sistemas.

En cuanto a la frecuencia de actualización de los recursos expuestos por este medio, dependerá de la configuración específica del repositorio y de cómo se implementa OAI-PMH. En general, el protocolo permite a los usuarios realizar solicitudes para obtener metadatos actualizados, lo que significa que la frecuencia de actualización puede ser ajustada según las necesidades del usuario y del repositorio. La recuperación de recursos desde los repositorios digitales puede realizarse siguiendo distintas estrategias. Una de ellas podría ser la implementación de un cosechador OAI-PMH por parte del sitio web, que recupere periódicamente registros desde los repositorios, los almacene

localmente, y luego muestre subconjuntos de estos recursos en las distintas secciones del sitio web.

El protocolo OAI-PMH permite la definición de diferentes contextos, y para cada contexto se definen los criterios de filtrado, esto es, qué elementos forman parte del contexto (por ejemplo, por tipología documental o colección de origen) y el formato de metadatos con el que se exponen los recursos (Dublin Core o MARC21). Estos contextos pueden definirse tanto estática como dinámicamente (De Giusti *et al.*, 2014; Mazurek & Werla, 2008). Cabe aquí destacar que el uso de un modelo estandarizado de representación de recursos resulta muy conveniente, ya que simplifica de manera considerable los algoritmos de procesamiento de los registros recuperados. Sin embargo, el uso de este protocolo no parece una solución ideal en caso de necesitar el recurso preservado, ya que fue pensado para realizar cosechas masivas periódicas, y requeriría que los sitios web institucionales gestionen un gran volumen de recursos en sus bases de datos propias, lo cual excede el propósito de un sitio web institucional tradicional. Asimismo, el uso del protocolo OAI-PMH, al no permitir consultas en vivo, no favorece las actualizaciones automáticas de los contenidos de los sitios web, y para mantener los sitios al día sería necesario realizar cosechas en intervalos de tiempo muy reducidos (diarias, o incluso más de una por día en los casos de instituciones con una gran producción propia) o bien configurar una tarea de indexación periódica en el servidor en función de la cantidad de recursos depositados en un día.

### **RSS, OpenSearch y API REST**

Para realizar consultas en vivo, una estrategia interesante es el uso del protocolo OpenSearch, basado en RSS/Atom. Este protocolo es soportado por un gran número de sistemas informáticos para repositorios digitales, como **DSpace**, **Open Preprint Systems (OPS)**, **EPrints** o **Digital Commons**. Asimismo, muchos repositorios digitales globales o temáticos también brindan servicios

de actualizaciones sobre RSS, como por ejemplo [Arxiv](#), [PubMed](#) y [bioRxiv](#). Por el lado de los CMS, un gran número de desarrollos incluyen o permiten incluir fácilmente servicios de recuperación y exposición de RSS de terceros, y de hecho existen experiencias en las que se utilizó el protocolo OpenSearch para exponer contenidos de repositorios digitales en sitios web institucionales (Villarreal *et al.*, 2017), con módulos específicos para Wordpress y Choique-CMS<sup>1</sup>. Por otro lado, a partir de la salida de la versión 7.0 del software DSpace, es también posible realizar consultas en vivo a repositorios con esta herramienta utilizando la nueva API REST de DSpace 7 (DSpace 7 REST API, 2023). A continuación, analizamos algunas ventajas y desventajas de ambas alternativas como soluciones para integrar recursos alojados en repositorios digitales con sitios web institucionales.

### **Ventajas y desventajas**

Por el lado de OpenSearch, la primera ventaja que podemos observar es que permite consultar y recuperar información utilizando el estándar RSS, basado en el protocolo HTTP, el cual se encuentra incorporado en una amplia variedad de herramientas actuales como Wordpress o Drupal, lo que permite que se puedan incorporar fácilmente las consultas en una gran cantidad de desarrollos. Otra de las ventajas que otorga OpenSearch es la simpleza y estandarización de la estructura de las respuestas que genera este protocolo. Esto hace que procesar la información generada a partir de una consulta sea relativamente sencillo, independientemente del repositorio al que se le realice.

Ya mencionadas las ventajas de este protocolo, es necesario revisar sus puntos débiles. Si bien la respuesta está estandarizada, ofrece una cantidad de metadatos limitada, por lo que la información que es posible extraer de cada

---

<sup>1</sup> El desarrollo de Choique CMS fue discontinuado en el año 2016. Su código fuente se encuentra disponible en el repositorio GitHub: <https://github.com/Desarrollo-CeSPI/choique>



ítem se encuentra acotada. Además, el protocolo RSS genera las respuestas en formato XML, mientras que un número cada vez mayor de sistemas están comenzando a utilizar JSON como formato de transporte de datos, gracias a su facilidad de procesamiento y menor tamaño requerido para representar información estructurada.

También es necesario analizar la construcción de consultas con OpenSearch. Para empezar, el formato de las consultas pueden variar en cada repositorio, ya que poseen muchos parámetros configurables. Por ejemplo, para consultar por autor en el repositorio SEDICI se usa la palabra clave *author*, pero en el repositorio institucional de CONICET se utiliza *dc.contributor.author*. Sumado a lo anterior, los parámetros OpenSearch disponibles en distintos repositorios digitales pueden encontrarse muy limitados, lo que en ocasiones no permite obtener el conjunto de datos deseado. Por ejemplo, si un repositorio organiza sus recursos en colecciones vinculadas a centros de investigación y no a tipologías documentales, y no permite filtrados por tipologías mediante OpenSearch, será muy complejo recuperar todos los recursos de un tipo particular (por ejemplo, libros) realizados en un centro de investigación (en términos de OpenSearch, filtrando por el *scope*).

En resumen, si bien el protocolo OpenSearch es fácil de utilizar, gracias a estar basado en RSS y a la estandarización y simpleza en sus metadatos, posee algunas desventajas a la hora de consultar distintos repositorios, ya que requiere configuraciones particulares para cada caso, y posee poca flexibilidad para parametrizar las consultas. Asimismo, este protocolo limita de manera considerable la información que es posible obtener de las fuentes.

Como alternativa al protocolo OpenSearch, es importante analizar las ventajas y desventajas del uso de la API REST incorporada en la versión de DSpace 7 (DSpace 7 REST Contract, 2023). Entre los principales puntos positivos se destaca la estandarización del formato de los requests HTTP para construir las consultas; al tratarse de una API REST, los endpoints a consultar y filtros a utilizar serán siempre los mismos entre los distintos repositorios

digitales, por lo que prácticamente no hace falta realizar configuraciones a la hora de construir las consultas. Además, esta API REST permite una gran cantidad de filtros y parámetros, lo cual brinda flexibilidad a la hora de obtener un conjunto de datos acotado y preciso de acuerdo con la información buscada. Por último, las respuestas que ofrece este mecanismo se encuentran en formato JSON, y poseen mucha más información sobre cada ítem en comparación con la que otorga las respuestas de consultas por OpenSearch.

En cuanto a los puntos negativos de la API REST, podemos destacar que las respuestas no poseen campos estandarizados por lo que procesarlas puede requerir configuración individual para cada repositorio. Además, en el contexto de análisis de este trabajo, los sistemas de gestión de contenidos no suelen estar integrados con API REST del mismo modo que lo hacen con OpenSearch, lo que requerirá esfuerzos adicionales de desarrollo a la hora utilizar este protocolo para recuperar recursos desde repositorios digitales y procesar las respuestas en formatos JSON.

## **Escenarios**

A continuación, se describirán algunos escenarios posibles en los que es dable aprovechar las capacidades de interoperabilidad para integrar repositorios digitales con sitios web institucionales. Si bien quedarán muchos escenarios por imaginar, el listado propuesto sirve para indagar sobre distintas formas de integración, y revisar requerimientos y limitaciones de las herramientas y tecnologías analizadas. Para cada escenario se describe su utilidad, y se incluyen ejemplos concretos sobre su uso o aplicabilidad en contextos reales. Asimismo, algunos escenarios incluyen también requerimientos técnicos particulares que deben cumplirse, o usos alternativos de acuerdo con las facilidades técnicas disponibles en cada contexto.

### ***Escenario 1: Exposición de reglamentaciones***

Las instituciones académicas pueden mostrar en sus sitios web institucionales las reglamentaciones, resoluciones y ordenanzas publicadas y almacenadas en su repositorio institucional. Para ello, el protocolo de interoperabilidad deberá permitir recuperar este tipo de recursos, ya sea a partir de su tipología documental o de la colección a la que pertenece. Esta información puede ser expuesta agrupada por año y por categoría, y ordenada por fecha de publicación. Por ejemplo, [este enlace](#) permite acceder a las resoluciones aprobadas por la Universidad Nacional de La Plata, durante el período 2010-2019, a través de su repositorio SEDICI. Utilizando el protocolo OpenSearch, puede recuperarse a través de [esta URL](#) de consulta.

De este modo, cada vez que algún área de la institución aprueba una nueva reglamentación y realice su depósito en el repositorio institucional, y que desde este espacio se completen los procesos de catalogación, descripción y preservación, se mostrará de inmediato en el sitio web de la institución. Esto resulta de gran utilidad para exponer en todo momento la información más actualizada vinculada a este tipo de documentos, incluso cuando su generación se realiza de manera descentralizada.

### ***Escenario 2: Producción científica de centros de investigación***

Los centros de investigación pueden exponer en sus sitios web la producción de artículos científicos, trabajos en congresos y proyectos de investigación. Por ejemplo, el Centro de Geología Aplicada, Agua y Medioambiente (CGAMA) de la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC) y de la Universidad Nacional del Sur (UNS) expone en su sitio web todos los artículos publicados por sus investigadores y alojados en el repositorio institucional CIC DIGITAL. En [este enlace](#) puede observarse esta integración. Para obtener estos recursos, el sitio web del CGAMA puede realizar [esta consulta](#) al repositorio CIC-DIGITAL por medio de la API REST de DSpace 7.

### **Escenario 3: Lecturas recomendadas**

En el contexto de entornos de enseñanza virtuales o LMS, así como también en sitios web de cátedras y departamentos, los docentes pueden incluir listados de recursos de la propia institución que los alumnos podrían aprovechar para enriquecer sus conocimientos y fortalecer los conceptos vistos durante las clases. Aquellas instituciones que posean Repositorios de Objetos de Aprendizaje pueden también aprovechar estas plataformas para incorporar recursos en sus propios LMS (Kalou *et al.*, 2008). En este escenario, resulta esencial contar, del lado del repositorio institucional, con capacidades de filtrado de documentos por área temática, a fin de ofrecer a los alumnos recursos académicos elaborados por cualquier investigador, independientemente del departamento o centro de investigación en el que realice sus actividades. Por ejemplo, el repositorio SEDICI permite consultas a través de OpenSearch filtrando por *subject*, como puede verse en [este enlace](#).

En el caso de repositorios digitales que funcionen sobre DSpace 7, la consulta puede resolverse consultando directamente los datos expuestos a través del módulo de búsqueda “discovery”. En el siguiente ejemplo se recuperan los recursos cuya palabra clave sea “Patrimonio” en el [repositorio CIC-DIGITAL](#).

### **Escenario 4: Producción de alumnos de un curso o carrera**

Los responsables del dictado de cursos y/o carreras pueden utilizar los repositorios de su institución para alojar y difundir trabajos de cursos, trabajos finales integradores de carreras, tesinas de grado y tesis de postgrado. Al hacer esto, podrán exponer en sus sitios web los trabajos realizados por sus propios alumnos, lo cual puede servir como ejemplo a otros alumnos para desarrollar sus propios trabajos y también como base para continuar y extender trabajos previos. Esto resulta en particular interesante para las áreas dedicadas a coordinar carreras de postgrado, ya que podrán mostrar a sus alumnos actuales, así como también a personas interesadas en realizar alguna carrera,

las tesis de doctorado y maestría que han realizado otros alumnos. Esta funcionalidad también puede resultar útil para organismos financiadores de becas doctorales, ya que podrían mostrar en sus propios sitios web las tesis ya aprobadas por sus becarios. Por ejemplo, [este enlace](#) permite acceder a las tesis doctorales realizadas por alumnos de la Facultad de Informática de la UNLP desde el repositorio SEDICI, y [esta URL](#) permite recuperar esta información por OpenSearch.

Si se intenta obtener esta información desde el repositorio CONICET DIGITAL, no es posible obtener tesis doctorales de la misma Facultad, pero sí es posible obtener todas las tesis doctorales a través de [esta URL](#).

Sin embargo, dado que la implementación de OpenSearch de CONICET DIGITAL no soporta filtrado por tipo de documento, no es posible obtener el mismo conjunto de recursos mediante este protocolo. En estos casos, la falta de estandarización a la hora de implementar interfaces OpenSearch en los repositorios digitales presenta un problema para brindar un mismo servicio de recuperación de datos desde múltiples repositorios.

### ***Escenario 5: Gobierno abierto y transparencia***

Las instituciones financiadas con fondos públicos deben promover la transparencia en todas sus acciones y brindar a la población general información sobre contrataciones, licitaciones, inversiones y compras en general. En el caso de Argentina, se fomentan y generan espacios de interacción entre la sociedad y el Estado para articular la colaboración y participación ciudadana en las distintas etapas del ciclo de las políticas públicas, y esto ha llevado a la implementación de distintos planes vinculados a Datos Abiertos (Datos Abiertos, 2023) y a gobierno abierto (Quinto Plan de Acción Nacional de Gobierno Abierto (2022-2024), 2023). Dichos planes involucran también a las instituciones del Sistema Nacional de Ciencia y Tecnología (SNCT), que pueden aprovechar sus repositorios digitales para recuperar los documentos vinculados a concursos de incorporación de

personal o de cambios de categorías, reportes de licitaciones, informes de compras y contrataciones a terceros. Por ejemplo, el repositorio SEDICI posee una colección dedicada a [Documentos Institucionales de la UNLP](#), desde la cual es posible acceder a subcolecciones y comunidades dedicadas a auditorias, planes estratégicos, licitaciones y concursos. En este sentido, es posible recuperar las licitaciones a través de la interfaz OpenSearch por medio de [esta URL](#).

### **Escenario 6: Perfiles de autores**

Un uso muy interesante de las facilidades de interoperabilidad entre repositorios digitales y sitios web institucionales es la posibilidad de generar perfiles institucionales de investigadores, docentes, becarios y autoridades en general, que expongan toda su producción científica y académica y que se actualicen de manera automática. En este caso, cada perfil puede incluso mostrar la producción alojada en diferentes repositorios, tanto institucionales, lo que resulta muy conveniente para autores de múltiple dependencia, como también globales o temáticos, como por ejemplo arXiv o PubMed.

Por ejemplo, el sitio web del Instituto Argentino de Radioastronomía (IAR) posee páginas dedicadas a sus miembros, y en dichas páginas puede verse la producción científica recuperada directamente desde el repositorio CONICET DIGITAL. Esto puede observarse en la página del Dr. Gustavo Romero, disponible en [este enlace](#).

Para casos de autores con múltiples dependencias, puede tomarse como ejemplo los perfiles de investigadores del Centro de Servicios en Gestión de Información (CESGI) de la CIC. [Aquí](#) puede observarse el perfil de la Dra. Marisa R. De Giusti, que incluye publicaciones recuperadas desde tres repositorios institucionales: SEDICI (UNLP) CIC-DIGITAL (CIC) y CONICET-DIGITAL (CONICET).

### **Escenario 7: Sistemas de agregación de recursos académicos**

Los sistemas de agregación de recursos académicos tienen como objetivo recopilar items desde distintos repositorios y agruparlos en un mismo conjunto; son sistemas que manejan grandes volúmenes de información y ésta suele ser actualizada periódicamente; un ejemplo de esto es el **Sistema Nacional de Repositorios Digitales (SNRD)**, el cual recopila recursos desde múltiples repositorios institucionales de Argentina, lo que permite al Ministerio de Ciencia, Tecnología e Innovación (MINCyT) mostrar desde un único espacio virtual la producción académica argentina al mundo. Asimismo, los recursos recuperados por el SNRD pueden ser cosechados por LA Referencia utilizando el mismo protocolo e integrándolos con recursos provenientes de otros países, lo que le permite generar servicios de valor agregado sobre un conjunto de recursos mucho más amplio. En este tipo de sistemas, la utilización del protocolo OAI-PMH resulta de gran utilidad, especialmente gracias al grado de estandarización de metadatos, las directrices y acuerdos generados entre las partes, y la facilidad de ejecutar cosechas masivas de recursos.

### **Análisis y conclusiones**

Los repositorios digitales proveen diferentes mecanismos para recuperar contenidos y exponerlos en sitios web institucionales. Entre estos mecanismos se destaca el protocolo OAI-PMH, como protocolo adecuado para la recuperación de grandes volúmenes de información de manera periódica, y los protocolos OpenSearch y API REST, como protocolos para la recuperación de volúmenes de información en vivo y más acotados, utilizando filtros. Si bien el protocolo OAI-PMH posee una amplia madurez y se encuentra muy afianzado en el ámbito de los repositorios digitales, su integración en sitios web institucionales resulta algo compleja debido a la poca integración nativa que por lo general ofrecen los CMS con este protocolo, y parece más adecuado para situaciones donde sea necesario mostrar una gran cantidad de recursos y

generar servicios como búsqueda y filtrado sobre ellos. Por otro lado, alternativas para consultas en vivo como el protocolo OpenSearch y la API REST de DSpace 7 parecen convenientes a la hora de recuperar conjuntos de recursos más acotados, basados en filtros que pueden ser más dinámicos y combinables. En este sentido, el protocolo OpenSearch resulta de momento más simple de integrar con sitios web institucionales implementados con gestores de contenidos como Wordpress o Drupal, gracias a estar basado en el estándar RSS, y a que estos sistemas ya incorporan o permiten incorporar muy fácilmente soporte para la recuperación y exposición de información a través de este protocolo; sin embargo, el protocolo RSS presenta importantes limitaciones en cuanto a la cantidad y a la estructura de información que permite recuperar y, por lo tanto, en cuanto al tipo de servicios que es posible brindar con su uso. Por otro lado, el uso de la API REST ofrece una mayor flexibilidad a la hora de construir consultas y permite recuperar una mayor cantidad de información de cada recurso; sin embargo, al tratarse de un desarrollo relativamente nuevo, aún no ha sido adoptado por un gran número de repositorios digitales y quizás su adopción masiva demore algunos años más. Asimismo, su integración con los CMS más utilizados hoy en día resulta algo más compleja, comparada con OpenSearch, ya que requiere implementar módulos *ad hoc* para generar y ejecutar los requests a la API REST, y procesar los resultados en formato JSON antes de presentarlos al usuario final.

## Bibliografía

- DE GIUSTI, M. R., ADORNO, F. G., LIRA, A. J. (2014). Repositorios DSpace con múltiples contextos OAI-PMH. En: IV Conferência Internacional sobre Bibliotecas e Repositórios Digitais (BIREDIAL) y IX Simpósio Internacional de Bibliotecas Digitais (SIBD) (Porto Alegre, Brasil, 2014). <<http://sedici.unlp.edu.ar/handle/10915/44572>>
- DE GIUSTI, M. R., VILLARREAL, G. L., TERRUZZI, F. A., OVIEDO, N. F., LIRA, A. J. (2013). Interoperabilidad entre el Repositorio Institucional y servicios en línea en la Universidad Nacional de La Plata. En: PKP International



- Scholarly Publishing Conferences (Mexico, 2013).  
<<http://sedici.unlp.edu.ar/handle/10915/27406>>
- DSPACE 7. (2023). REST API. Lyris.org.  
<<https://wiki.lyris.org/display/DSPACE/Dspace+7+REST+API>>
- DSPACE 7. (2023). REST Contract. GitHub.com.  
<<https://github.com/Dspace/RestContract/>>
- KALOU, A.K., KOUTSOMITROPOULOS, D.A., SOLOMOU, G.D., BOTSIOS, S.D. (2015). Metadata Interoperability and Ingestion of Learning Resources into a Modern LMS. In: Garoufallou, E., Hartley, R., Gaitanou, P. (eds) Metadata and Semantics Research. MTSR 2015. Communications in Computer and Information Science, vol. 544. Springer, Cham. DOI:  
<[https://doi.org/10.1007/978-3-319-24129-6\\_15](https://doi.org/10.1007/978-3-319-24129-6_15)>
- MAZUREK, C., & WERLA, M. (2008). Extending OAI-PMH protocol with dynamic sets definitions using CQL language. En: Conference proceedings of IADIS Information Systems, Algarve, Portugal (April 2008). IADIS Publications.
- OPENSOURCE CMS. (2023). Open-source CMS Market Share.  
<<https://www.opensourcecms.com/cms-market-share/>>
- REPÚBLICA ARGENTINA. (2023). Datos Abiertos.  
<<https://www.argentina.gob.ar/jefatura/innovacion-publica/servicios-y-pais-digital/gobierno-abierto/datos-abiertos>>
- REPÚBLICA ARGENTINA. (2023). Quinto Plan de Acción Nacional de Gobierno Abierto (2022-2024).  
<<https://www.argentina.gob.ar/jefatura/innovacion-publica/servicios-y-pais-digital/gobierno-abierto/quinto-plan-de-accion-nacional>>
- VILLARREAL, G. L., SALAMONE LACUNZA, P., VILA, M. M., DE GIUSTI, M. R., & MANZUR, E. (2017). A Simple Method for Exposing Repository Content on Institutional Websites. Open Repositories. En Open Repositories 2017 (Queensland, Australia, 2017).  
<<http://sedici.unlp.edu.ar/handle/10915/60507>>
- WPBEGINNER. (2023). 2023's CMS Market Share Report – Latest Trends and Usage Stats. WPbeginner.com.  
<<https://www.wpbeginner.com/research/cms-market-share-report-latest-trends-and-usage-stats/>>

# Interface do gestor: uma dashboard para gestores de repositórios digitais

## **JULIANA ARAUJO GOMES DE SOUSA**

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA

[julianasousa@ibict.br](mailto:julianasousa@ibict.br)

## **LAUTARO MATAS**

LA REFERENCIA

[lmatas@gmail.com](mailto:lmatas@gmail.com)

## **VIVIAN SANTOS SILVA**

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA

[vivian.ss@gmail.com](mailto:vivian.ss@gmail.com)

## **WASHINGTON R. DE CARVALHO SEGUNDO**

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA

[washingtonsegundo@ibict.br](mailto:washingtonsegundo@ibict.br)

## **JUAN SEBASTIAN MANITTA**

LA REFERENCIA

[manittajuan@gmail.com](mailto:manittajuan@gmail.com)

### **EIXO TEMÁTICO**

Infraestructura tecnológica

### **RESUMEN**

O Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) em parceria com a LA Referencia desenvolveu uma aplicação para visualização de dados que permitisse aos gestores de repositórios digitais ter acesso às métricas de coleta de dados realizadas. Para o

desenvolvimento da ferramenta utilizou-se de três tecnologias distintas, a primeira é o Software da Plataforma LA Referencia, que tem como função coletar, validar e enriquecer os metadados, já o keycloak é responsável pelo processo de autenticação, gerenciamento de usuários e apresentação das informações relativas aos repositórios e o Angular que utiliza-se das duas tecnologias anteriores para apresentar as métricas geradas por meio do processo de coleta. Este trabalho se concentra no Eixo 4 (Infraestrutura tecnológica). O objetivo é apresentar a interface do gestor, seu desenvolvimento, tecnologias utilizadas e suas funcionalidades. A metodologia utilizada é quali-quantitativa com natureza aplicada, pois tem a finalidade de solucionar de forma prática problemas específicos. Apesar da ferramenta ainda não estar em produção, identificou-se que ela será um importante aliado para auxiliar os gestores de repositórios a ter autonomia na realização da curadoria de metadados fazendo com que a qualidade dos dados coletados aumente de forma a gerar impacto direto na representação, recuperação e disseminação da informação científica.

#### **PALAVRAS CHAVE**

Visualização de dados; repositórios digitais; interface do gestor.

### **Introdução**

No contexto dos portais agregadores de informação científica, o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) em parceria com a Red de repositorios de acceso abierto a la ciencia (LA Referencia) desenvolveu um software que permitirá aos gestores de repositórios terem acesso ilimitado e irrestrito aos diagnósticos das coletas entre repositórios digitais e portais agregadores.

O desenvolvimento desse software denomina-se “interface do gestor” e ele permitirá que todas as instituições participantes dos 12 nós da rede LA Referencia possam ter acesso ao sistema de diagnóstico de coleta, que antes só era acessível a instituição responsável por cada nó em um país. Por exemplo, no Brasil apenas os responsáveis pela Biblioteca Digital Brasileira de Teses e Dissertações (BDTD) e pelo Portal brasileiro de publicações e dados científicos em acesso aberto (Oasisbr) tinham acesso a tela de análise dos

dados coletados das fontes de informação associadas à BDTD e ao Portal Oasisbr.

Diante disso, e com o objetivo de oferecer maior autonomia aos gestores de repositórios digitais que fazem parte de um dos doze nós<sup>1</sup> vinculados a Rede LA Referencia, desenvolveu-se um dashboard que possibilita a visualização dos dados gerados por meio da coleta dos repositórios. Portanto, os gestores passarão a ter acesso a quantidade de registros coletados, quantos desses registros foram validados, quantos foram invalidados e o porque o registro foi considerado inválido.

Com isso, o objetivo deste trabalho é apresentar as tecnologias utilizadas para o desenvolvimento da ferramenta, bem como apresentar as funcionalidades que serão disponibilizadas nesta primeira versão aos gestores de repositórios.

Para compreender como os dados são aglutinados, transformados, validados e apresentados em uma dashboard, foram utilizadas três ferramentas distintas e com funções específicas dentro da arquitetura da aplicação. Em primeiro plano, tem-se a plataforma de software LA Referencia, que valida cada um dos registros previamente coletados. O processo de validação tem como resultado uma série de métricas e informações que são armazenadas em um núcleo **SOLR**.

Em segunda camada tem-se a interface da dashboard, que é uma aplicação desenvolvida com o framework **Angular** que apresenta diferentes relatórios de validação que são baseados por meio de uma consulta realizada a uma Application Programming Interface (API), desenvolvida especificamente para

---

<sup>1</sup> Países que integram os 12 nós da Rede LA Referencia: Argentina, Chile, Brasil, Colômbia, Costa Rica, Equador, El Salvador, México, Espanha, Panamá, Peru e Uruguai.

consultar os dados gerados pelo processo de coleta e validação que é realizado pela plataforma LA Referencia.

A API da dashbord implementa funções de autorização para acesso de distintos usuários e/ou distintos grupos de repositórios. Para isso utilizou-se a ferramenta Keycloak, a qual é a terceira camada da aplicação, que é um provedor de identidades que possibilita designar níveis de autorização a cada perfil de usuário que é criado. Com isso um usuário com perfil de administrador pode agregar outros usuários e também ter administração de repositórios distintos. Ademais, a API permite múltiplos pontos de acesso aos relatórios de validação e histórico de coletas de um repositório cadastrado.

### **Procedimentos metodológicos**

Para a construção metodológica deste trabalho utilizou-se uma abordagem qualitativa de natureza aplicada. Esse método se justifica, pois a abordagem qualitativa foi utilizada na definição e realização de entrevistas guiadas e aplicada pois ao final do estudo tem-se como produto um novo serviço que objetiva atender as necessidades identificadas por meio das entrevistas realizadas.

As entrevistas foram realizadas em 11/05/2020, por meio da plataforma de reunião online Google Meet, com quatro gestoras de repositórios digitais brasileiros que fazem parte tanto da rede da BDTD quanto do Portal Oasisbr. A entrevista foi composta de uma única pergunta que foi “qual a necessidade de informação que elas, enquanto gestoras de repositórios, tinham em relação ao processo de coleta de dados?”

Por meio das respostas obtidas, a equipe pode traçar uma direção para atender às principais demandas apontadas pelas entrevistadas. Como se trata de uma metodologia qualitativa de natureza aplicada, primeiramente iremos apresentar os resultados obtidos por meio da entrevista e em sequência os

passos para o desenvolvimento da dashboard, a qual foi guiada pelas entrevistas.

## **Resultados**

Por meio de entrevista realizada em 11/05/2020 com 04 gestoras de repositórios, observou-se que os tópicos citados foram:

- Ter acesso aos dados quantitativos de uma coleta: quantos documentos foram coletados, quantos foram validados, quantos foram invalidados, quantos documentos novos são coletados a cada mês.
- Verificar quais as tipologias documentais estão sendo coletadas.
- Exportar relatórios de coletas de maneira personalizada.
- Para atender as necessidades observadas nas entrevistas, a equipe de desenvolvimento se baseou nos tópicos acima para permitir que a dashboard tenha funcionalidades desejadas. As funcionalidades serão melhores descritas no tópico sobre a interface do gestor.

## **Software da Plataforma LA Referencia**

O software da Plataforma LA Referencia está em sua quarta versão e é desenvolvido em linguagem Java, sob o framework Spring e o banco de dados PostgreSQL. A plataforma é responsável por realizar a coleta dos metadados via protocolo Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) das bibliotecas e repositórios digitais, além de realizar o enriquecimento e a validação dos metadados coletados. A plataforma organiza os dados cadastrais de cada instituição participante da Rede, tais como: nome da instituição, nome do repositório, se a rede é visível ao público, natureza administrativa da instituição, telefone, e-mail, tipo da fonte de coleta, tipos de

documentos a serem coletados, qual o software utilizado, sigla da instituição, Uniform Resource Locator (URL) e a URL URL OAI-PMH.

Pode-se dizer que dentro do formulário de cadastro de coleta de uma instituição a informação mais importante é a URL OAI-PMH, pois é por meio dela que os metadados serão coletados, transformados e validados. Cada uma das três etapas são essenciais para que os registros de uma instituição sejam apresentados no próprio portal da LA Referencia, que atualmente agrega mais 4 milhões de documentos<sup>2</sup> científicos em acesso aberto, bem como para apresentar os registros nos agregadores de cada um dos doze países que fazem uso da tecnologia.

Para unificar diversas fontes de informação distribuídas em diferentes contextos tecnológicos e organizacionais, foi necessário definir um padrão de metadados para que, independente da fonte de informação, os dados estivessem padronizados. Além da definição de um padrão comum de metadados, foi necessário definir quais destes metadados seriam de preenchimento obrigatório. Para isso utilizou-se as diretrizes Open Access Infrastructure for Research in Europe (OpenAIRE) para repositórios de literatura, institucionais e temáticos e também as [diretrizes OpenAIRE](#) para [arquivos de dados](#), a ser utilizados para os casos de Repositórios de dados de pesquisa. De acordo com o próprio guia

As [diretrizes OpenAIRE](#) fornecem orientação científica para que gestores de repositórios possam definir e implementar suas políticas de gestão de dados local...além do mais os gestores de repositórios não estarão apenas permitindo que autores, que depositam publicações em seu repositório, cumpram com os

---

<sup>2</sup> Os documentos coletados são artigos, teses, dissertações, livro, capítulo de livro, dados de pesquisa e relatórios.

requisitos de Acesso Aberto da Comissão Europeia, mas também e, eventualmente os requisitos de outros financiadores.

O impacto direto no uso das diretrizes OpenAIRE é que dentro de suas recomendações é definido um conjunto de requisitos para o preenchimento de determinados metadados. Essas recomendações atestam ainda quais metadados são de preenchimento obrigatório, obrigatório se aplicável, recomendado e opcional.

De acordo com essas determinações, a plataforma de software LA Referencia optou por utilizar como regra obrigatória apenas os metadados definidos como de uso obrigatório. É importante explicitar que o preenchimento em branco do metadado não deve ser dado como válido. Portanto, os metadados de preenchimento obrigatório para validação são: título, autor, data de publicação, tipo de documento, direitos de acesso e identificador persistente. É importante destacar que dentro do sistema de coleta e validação dos registros cada um desses metadados é definido como uma regra de validação.

Nos campos tipo de documento e direitos de acesso, o software utiliza o vocabulário controlado da Confederation of Open Access Repositories (COAR) . O vocabulário controlado para tipo de documento aceito no processo de validação é o definido pelo OpenAIRE Guideline versões 3 e 4 e também pelos vocabulários COAR, além de utilizar o vocabulário controlado da COAR para definir a tipologia documental, utiliza-se também para determinar quais são os direitos de acesso ao documento.

Portanto, para que os metadados que representam a informação de tipologia documental e direitos de acesso sejam sancionados no processo de validação a informação preenchida no campo deve estar de acordo com o que determina o vocabulário controlado citado.



No entanto, notou-se que muitas instituições não faziam uso dos vocabulários controlados para o preenchimento das informações, para que isso não causasse um impacto direto na validação dos itens coletados os registros

Até aqui apresentamos como é realizado o processo para coleta e validação dos dados, porém é necessário discorrer sobre o processo de transformação ou enriquecimento dos metadados.

De acordo com Carvalho et al (2021, p.2) “o software LA Referencia também tem uma função de enriquecimento e curadoria dos dados” e esse processo é feito durante a execução da transformação dos dados coletados, ou seja, o processo de transformação nada mais é do que um ato de padronização de informações semelhantes. Exemplo de um processo de transformação ocorre para o campo de tipo de documento. Como mostrado acima, para ser validado o campo deve ser preenchido de acordo com as normas do vocabulário controlado indicado. Caso isso não ocorra, o registro será invalidado e por seguinte não será apresentado na ferramenta de busca dos portais agregadores. Para dirimir esses problemas, criou-se o processo de transformação, que traduz informações despadronizadas para o padrão definido (ver Figura 1).

Traducción

Buscar

artículo de revista

Reemplazo

info:eu-repo/semantics/article

Traducción

Buscar

Artículo Internacional

Reemplazo

info:eu-repo/semantics/article

Traducción

Buscar

Artículo revisado por pares

Reemplazo

info:eu-repo/semantics/article

FIGURA 1. Processo de transformação para o campo de tipo de documento

Fonte: PrintScreen de uma das regras de transformação do La Referencia Harvester

A figura 1 apresenta um exemplo de um processo de transformação, o qual funciona da seguinte forma:

- No campo “Buscar” deve-se inserir a informação que a instituição preenche originalmente no campo.
- No campo “Reemplazo” deve-se inserir a informação correspondente no vocabulário controlado.

Realizado os passos acima, o sistema automaticamente substituirá a informação preenchida pela instituição e irá inserir a instituição definida por padrão. Dessa forma, não só o campo de tipologia documental, que os campos têm suas informações padronizadas e tratadas diretamente no software da Plataforma LA Referencia.

Ao final do processo de coleta, transformação e validação dos registros são gerados dados quantitativos referente a cada uma das regras de coleta. No entanto, essas métricas só são geradas se não ocorrer falha em nenhuma das três etapas.

Antes do desenvolvimento da dashboard, somente pessoas com acesso ao software LA Referencia podiam ter acesso aos dados. No entanto, com o lançamento da interface do gestor, todas as instituições passarão a ter acesso a esses dados. Para tornar isso possível, foi necessário utilizar uma ferramenta que permitisse o gerenciamento de usuários, e para esse propósito, optou-se pelo Keycloak.

## **Keycloak**

Os usuários devem estar autenticados para ter acesso à Interface do Gestor. Para isso foi utilizado o Keycloak, uma ferramenta de autenticação e autorização que permite a definição de permissões de acesso de forma flexível, no nível de usuário, grupos de usuários, ou papéis (roles), ou ainda através de regras definidas de acordo com as necessidades da aplicação. No contexto da Interface do Gestor, a permissão de acesso a repositórios é baseada em uma regra que determina que um usuário comum (não administrador) deve ser membro de um grupo identificado pelo acrônimo de um repositório para ter acesso às informações deste repositório. Usuários administradores, por outro lado, têm acesso irrestrito a todos os repositórios. O Keycloak se encarrega de verificar as informações do usuário autenticado, aplicar as regras convenientes e decidir se concede ou não o acesso.

A Ferramenta Keycloak possui uma interface web de administração (para cadastro de aplicações, inclusão de usuários, definição de permissões, etc.), mas também permite a administração de forma programática, permitindo que a criação e atualização de usuários, individualmente ou em massa, possa ser feita diretamente dentro da Interface do Gestor da Biblioteca/Repositório

Digital. Para isso, utilizando a API de administração disponibilizada pelo Keycloak, a API REST da plataforma LA Referencia foi estendida para prover também funcionalidades de gestão de usuários. Dentro da Interface do Gestor, um módulo de administração de usuários, visível apenas para usuários com perfil de administrador, permite a criação, edição e remoção de usuários e grupos, e inclusão e exclusão de usuários em grupos (o que irá determinar suas permissões de acesso, ou seja, que repositórios poderá visualizar). Além disso, também é possível carregar informações de uma lista de usuários de um arquivo Comma-separated values (CSV) e sincronizar a lista de grupos com a lista de repositórios na base de dados da plataforma LA Referencia. Desta forma, qualquer administrador na Interface do Gestor pode gerenciar usuários de forma simples e transparente, sem precisar lidar diretamente com o Keycloak.

### **A interface do gestor de repositórios**

Para o desenvolvimento da interface de visualização dos dados foram utilizados conceitos de Single-Page Application (SPA) com a utilização do framework Angular em sua 15ª versão, dando ênfase em interfaces de usuário interativas, contando com o poder de tipagem do typescript e facilidade de reuso de componentes e acoplamento em módulos. Uma API REST, desenvolvida como um módulo da plataforma LA Referencia com auxílio de Spring Boot, é consumida pela dashboard para recuperação dos dados relativos ao histórico de coletas e de validação dos registros.

Esta API permite a recuperação da lista completa de repositórios coletados, das informações de um repositório específico, da lista de coletas de um determinado repositório, incluindo data e horário de início e fim da coleta e quantidade de registros coletados, e dos resultados da validação para cada coleta, incluindo o número de registros válidos e inválidos, e, para os registros inválidos, quais regras foram violadas. A internacionalização, por meio do

Angular Internationalization (i18n), foi utilizada para a tradução da interface em diferentes idiomas e a formatação de dados para locais específicos.

O desenvolvimento da interface do gestor foi pensado para permitir que os gestores tenham acesso aos processos de coletas e validação dos repositórios. A pertinência do desenvolvimento dessa ferramenta deu-se ao entrevistar alguns gestores para entender qual a necessidade de informação que eles tinham que fossem relacionadas ao processo de coleta. As principais dúvidas estão associadas aos dados quantitativos, principalmente, quando o número de documentos apresentados no repositório era sumariamente maior do que o quantitativo apresentado nos portais agregadores. Essas dúvidas cotidianas dos gestores ocasionaram demandas significativas aos responsáveis pelos nós para analisar o diagnóstico de cada instituição e respondê-las de forma que eles pudessem compreender o problema e trabalhar na resolução, visto que o interesse em coletar e apresentar dados de qualidade é de ambos os lados.

Isto posto, deu-se início ao processo de criação da interface, tendo em mente que os principais usuários dessa ferramenta serão bibliotecários ou cientistas da informação, de modo que os dados deveriam ser expostos de maneira compreensível para que estes profissionais tenham autonomia em entender os dados apresentados na dashboard.

Para que isso fosse possível, foi necessário realizar um trabalho de verificação das informações apresentadas e criar novos textos para que os dados apresentados ficassem claros para o grupo de usuários.

Ao acessar a dashboard a primeira tela que o gestor terá acesso apresentará um o histórico de coletas, ou seja, a cada período é realizada uma coleta completa do repositório, cada uma dessas coletas gera um identificador único. Portanto, a primeira métrica que o usuário da interface do gestor irá visualizar será o histórico de coletas, onde será apresentado o identificador único da coleta, a quantidade de registros coletados, a quantidade de registros válidos e

inválidos. Por meio do histórico de coletas, o gestor poderá verificar quantos documentos foram coletados a cada mês ou a cada coleta. Além disso, será possível exportar esses dados em formato de tabela. Conforme apresenta a figura abaixo.

board / Brasil Powered by LA Referencia

> Harvesting table - Read more

ID ↑↓	Harvested Record ↑↓	Valid records ↑↓	Invalid records ↑↓	Start ↑↓	End ↑↓	
2,865	2,128,408	2,127,215	1,193	2/16/23, 10:30 AM	2/16/23, 7:57 AM	
2,688	1,583,624	1,583,542	82	8/16/22, 12:03 PM	8/17/22, 2:43 AM	
2,600	1,550,003	1,549,921	82	7/8/22, 1:34 PM	7/9/22, 8:35 AM	
2,566	1,467,332	1,467,065	267	5/2/22, 3:10 PM	5/4/22, 12:03 PM	
2,439	1,385,033	1,377,000	8,033	7/26/21, 8:51 AM	7/26/21, 8:35 PM	
2,385	1,490,108	1,483,691	6,417	5/20/21, 4:55 PM	5/21/21, 8:54 AM	
2,330	1,545,248	1,544,423	825	3/20/21, 9:31 AM	3/20/21, 9:54 PM	

FIGURA 2. Apresentação do histórico de coletas de uma instituição na interface do gestor

Fonte: PrintScreen da aplicação no sistema desenvolvido para a Interface do Gestor

Os dados da figura 2 também poderão ser visualizados em forma de gráfico, onde cada coleta será representada por uma barra com a data e a quantidade de registros coletados, validados e invalidados. Esta visualização está representada na Figura 3.



FIGURA 3. Apresentação do histórico de coletas de uma instituição na interface do gestor

Fonte: PrintScreen da aplicação no sistema desenvolvido para a Interface do Gestor

Também é possível visualizar as métricas relativas aos registros válidos e inválidos separando os dados quantitativos por tipo de regra, ou seja, ao coletar um repositório é possível visualizar quantos dos registros coletados foram validados e invalidados para o campo de autoria, título, tipo de documento, data de publicação, direitos de acesso e identificador persistente. Para isso, os dados são apresentados em uma tabela em que cada linha corresponde a uma regra, conforme apresenta a Figura 3:

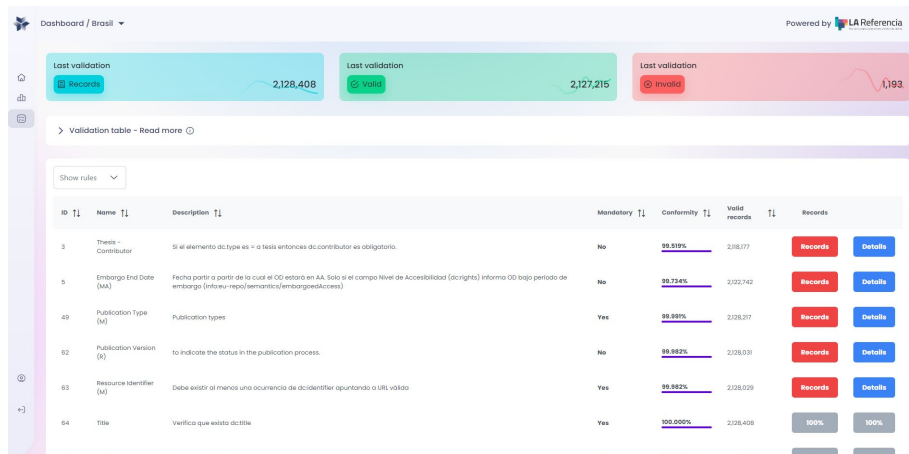


FIGURA 4. Tela de diagnóstico por tipo de regra

Fonte: PrintScreen da aplicação no sistema desenvolvido para a Interface do Gestor

Há um contador de ocorrências de termos para cada regra aplicada, possibilitando uma análise de similaridade de ocorrências. Além disso, é possível ter acesso individual sobre cada registro da coleta, analisar cada regra aplicada, verificar cada ocorrência examinada, e visualizar o XML do registro transformado.



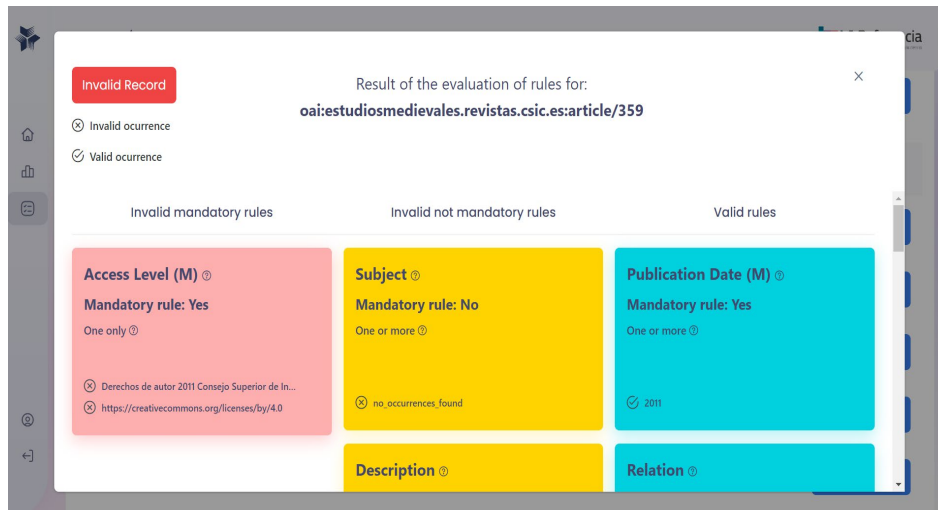


FIGURA 5. Aplicação das regras ao registro e acesso ao XML transformado

Fonte: PrintScreen da aplicação no sistema desenvolvido para a Interface do Gestor<sup>3</sup>

As métricas acima serão disponibilizadas na primeira versão da interface do gestor. Apesar de ter informações claras sobre os dados apresentados, a equipe desenvolvedora entende que será necessário treinar os futuros usuários da ferramenta para que o objetivo principal da ferramenta possa ser alcançado.

### Considerações finais

O objetivo principal para o desenvolvimento da interface do gestor se concentrou em dar acesso aos relatórios de coleta de dados aos gestores dos repositórios digitais, no entanto, a tecnologia desenvolvida também auxilia os gestores a realizarem a curadoria dos metadados diretamente na fonte, sem que seja necessário utilizar a funcionalidade de transformação da Plataforma de software LA Referencia.

---

<sup>3</sup> As imagens e os dados apresentados nas figuras 2, 3, 4 e 5 foram retirados do ambiente de teste e no momento as definições de layout ainda não tinham sido aplicadas.

Esse processo de autonomia na realização da curadoria dos metadados impacta diretamente na qualidade dos dados coletados; na representação e recuperação da informação; na diminuição dos registros invalidados no processo de coleta e, principalmente, contribui de maneira indireta na disseminação, compreensão e uso das boas práticas para o compartilhamento da informação científica na web por meio de portais agregadores.

A interface do gestor já está em etapa de conclusão, tendo os seus módulos de integração já consolidados. O trabalho atual consiste em realizar ajustes no código da dashboard e também refinar os recursos relacionados à experiência do usuário e também ao design da aplicação.

No que se refere à confiabilidade dos dados apresentados na dashboard, durante o ano de 2021 foram realizados diversos testes de visualização dos dados que estavam sendo apresentados, portanto, neste quesito a ferramenta já está estável.

Diante disso, espera-se que ainda em 2023 a interface do gestor seja finalizada e distribuída para a implementação, gerência e uso dos doze nós integrantes da Rede LA Referencia.

## Bibliografía

- CARVALHO, J., MATAS, L., SEGUNDO, W., GRAÇA, P., & LOPES, P. (2021). Dos repositórios aos agregadores, o metamodelo de relações entre entidades: o caso LA Referencia e RCAAP. 11<sup>a</sup> Confoa. <<http://aleph.letras.up.pt/index.php/paginasueb/article/view/10244/9511>>
- CONFEDERATION OF OPEN ACCESS REPOSITORIES (COAR). (2022). Controlled vocabularies for repositories. <[https://vocabularies.coar-repositories.org/access\\_rights/](https://vocabularies.coar-repositories.org/access_rights/)>
- OPEN ACCESS INFRASTRUCTURE FOR RESEARCH IN EUROPE (OpenAIRE). (2018). OpenAIRE Guidelines for Literature Repository Managers v4. <<https://openaire-guidelines-for-literature-repository-managers.readthedocs.io/en/v4.0.0/>>

- OPEN ACCESS INFRASTRUCTURE FOR RESEARCH IN EUROPE (OpenAIRE). (2020). Diretrizes OpenAIRE para Repositórios de publicações científicas v4. <<https://livroaberto.ibict.br/handle/123456789/1087>>
- OPEN ACCESS INFRASTRUCTURE FOR RESEARCH IN EUROPE (OpenAIRE). OpenAIRE Guidelines for Data Archives. (2022). <<https://guidelines.openaire.eu/en/latest/data/index.html>>
- SCHMIDT GODOY, A. (1995). Pesquisa qualitativa. *Revista de Administração de Empresas*, 35(3), 20-29. <<https://www.scielo.br/j/rae/a/ZX4cTGrqYfVhr7LvVyDBgdb/?format=pdf&lang=pt>>

# Vitrinas del conocimiento: una oportunidad para la acción

## MARISOL GOYENECHÉ REINA

FUNDACIÓN UNIVERSITARIA DE CIENCIAS DE LA SALUD (FUCS)

[mgoyeneche@fucsalud.edu.co](mailto:mgoyeneche@fucsalud.edu.co)

## ANA LORENA NIÑO TÉLLEZ

FUNDACIÓN UNIVERSITARIA DE CIENCIAS DE LA SALUD (FUCS)

[bvscolombia@fucsalud.edu.co](mailto:bvscolombia@fucsalud.edu.co)

### EJE TEMÁTICO

Comunicación académica, científica y cultural en abierto

### RESUMEN

La Vitrina del Conocimiento es un recurso del modelo Biblioteca Virtual de Salud, que pretende dar a conocer documentos seleccionados y recursos de información sobre un tema de interés para la salud (BIREME, s.f.). Como un recurso innovador que apoya el proceso de planificación, identificación y diseño de herramientas para la gestión del conocimiento, presenta elementos diferenciales estructurados de acuerdo con su propia metodología, y cuenta con dos áreas principales: una de componentes, donde presentan elementos destacados de información y un área de contenido; ambos brindan acceso directo a documentos de calidad, elementos específicos para la temática e información en diferentes formatos, como videos, infografías, infometría o cualquier otro recurso que permita al investigador acceso rápido y fiable a la información seleccionada. El proceso de construcción de la vitrina requiere del trabajo interdisciplinario con expertos en el tema propuesto, quienes aportan sus conocimientos para la construcción y selección de contenidos de la herramienta; la metodología es de acceso abierto y la construcción voluntaria. La Biblioteca Virtual en Salud (BVS), es una estrategia que promueve el acceso abierto, arbitrado, equitativo y de calidad a la información; formada como una red de colaboración entre productores, intermediarios y usuarios de información, como una estrategia para América Latina y el Caribe. (Niño Téllez, Acercamiento a la visibilidad de la información en salud en Colombia y la Biblioteca Virtual de Salud, 2016).

### **PALABRAS CLAVES**

Ciencia de la información; biblioteca digital; gestión del conocimiento; procesamiento de la información. Information processing; information sciences; digital libraries; knowledge management.

### **Vitrinas de Conocimiento**

Las Vitrinas de Conocimiento son un recurso de la BVS que busca resaltar documentos y recursos de información seleccionados sobre un tema de salud relevante. Tiene la misma función que un escaparate en sí, llamando la atención sobre un conjunto seleccionado y representativo de contenidos relacionados a un tema, que puede estar asociado o no a un Portal de Instancias de la Red BVS.

Las Vitrinas de Conocimiento siguen una metodología propia, dando acceso directo a documentos indexados en las fuentes de información de la BVS, enlaces a estrategias de búsqueda predefinidas, fuentes de información externas, sitios y páginas fundamentales para el tema en cuestión, o incluso contenidos actualizados vía RSS (Really Simple Syndication) como blogs, boletines epidemiológicos, noticias y otros. Ofrece además información en diferentes formatos como videos, infografías, infometría o cualquier otro recurso que permita al investigador acceder de manera rápida y confiable a la información (Centro Latinoamericano y del Caribe de Información en Ciencias de la Salud - BIREME/OPS/OMS, 2020). Las vitrinas del conocimiento conforman una amplia colección de recursos propuestos y diseñados según las necesidades de las BVS regional, temáticas y nacionales.

### **Modelo BVS**

El modelo biblioteca virtual de salud (BVS) es una estrategia democratizada de gestión y acceso libre a la información para América Latina y el Caribe, la iniciativa surge en 1998 en el marco del IV Congreso Regional de Información

en Ciencias de la Salud reunido en San José de Costa Rica, donde se emite la declaración «Hacia la Biblioteca Virtual de Salud», con el compromiso de renovar la entrega y procesamiento de la información científico-técnica producida en la región, resultado de los cambios tecnológicos y la posibilidad de hacer uso de ellos para dar visibilidad a la información producida en cada país.

Conformada por redes de cooperación y estrategias que permiten acceder a la información de manera libre, oportuna, confiable y pertinente, la BVS es una propuesta que asegura respaldo y trascendencia a la visibilidad de la producción científica técnica de cada país, además de fortalecer y liderar el proceso del flujo de la información (recopilación, registro, control de calidad, preservación y disseminación) aumentando la visibilidad, accesibilidad, calidad, credibilidad, y el uso e impacto de la literatura científica y técnica producida en el país, y en la región del Caribe y latinoamericana.

### **Objetivos**

Dar a conocer la Vitrina del conocimiento como un recurso innovador del modelo Biblioteca Virtual de Salud – BVS, como una estrategia de difusión y gestión de información en salud.

### **Objetivos específicos**

- Proporcionar un recurso de divulgación para la descripción, gestión y recopilación de fuentes de información externa con indicación de sitios y páginas fundamentales para el tema en cuestión.
- Facilitar la comprensión de contenidos en áreas relacionadas en salud, para apoyar la toma de decisiones soportadas.

## Marco histórico

### BIREME

El Centro Especializado BIREME es una iniciativa de cooperación técnica de la OPS/OMS, con la finalidad de apoyar la gestión de información científica en salud de las Américas. Establecido en 1967 en Sao Paulo, Brasil, su trabajo está orientado en el desarrollo de capacidades e infraestructuras nacionales y regionales para la gestión, acceso, organización, indexación, preservación y difusión de la información, el conocimiento y la evidencia científica en concordancia con el estado de las metodologías de arte y tecnologías relacionadas.

Durante el establecimiento de BIREME, la cooperación técnica realizada por la OPS a través de BIREME ha logrado una notable contribución a la democratización de la información de calidad, esencial para el desarrollo de la salud y la inclusión social en esta región.

A través del tiempo la cooperación ha evolucionado con los modelos de gestión de la información y la comunicación científica, iniciando en 1967 como la Biblioteca Regional de Medicina, basada en las funciones esenciales de las bibliotecas médicas, incluida la formación en gestión y el funcionamiento de las bibliotecas y centros de documentación, desarrollo de colecciones locales, cooperación en el uso de las colecciones entre bibliotecas, servicios al usuario, búsquedas bibliográficas en MEDLINE y fotocopias de documentos.

Hasta finales de los años 80, el modelo de cooperación técnica se extiende, con la nueva función de control bibliográfico e indización de la literatura publicada en revistas científicas de salud en América Latina y el Caribe y, por otro lado, la ampliación de la cobertura temática para cubrir la totalidad del campo de ciencias de la salud. En 1979, BIREME lanzó el Index Medicus Latino-Americano (IMLA), con la indexación de cerca de 150 revistas y complementando por lo tanto MEDLINE, que indexa 44 títulos en América Latina y el Caribe. Con IMLA, BIREME comenzó su papel destacado en la

visibilidad regional e internacional para la producción científica y técnica en salud en América Latina y el Caribe. Esta expansión originó el cambio de nombre, de Biblioteca Regional de Medicina a Centro América Latina y el Caribe las Ciencias de la Información de salud, sin embargo, se mantiene la sigla BIREME. Durante este período, el IMLA evolucionó a la base de datos bibliográfica en Ciencias de la Salud de América Latina y el Caribe (LILACS), con el propósito de ampliar la cobertura de información en salud pública.

A finales de los años 80, BIREME promueve la descentralización, a nivel de país, las funciones de control bibliográfico de la literatura científica, los servicios de búsqueda en la literatura y la formación de los recursos humanos. Esta descentralización, estructura la instancia nacional, coordinada en cada país. Así mismo, BIREME promovió el uso de ordenadores en las bibliotecas, tanto para la producción descentralizada de la base de datos LILACS y para la investigación bibliográfica en CD-ROM y más tarde en línea. También en este período, se enriqueció las metodologías de gestión de la información científica con el lanzamiento y actualización anual del vocabulario en el Descriptor en Ciencias de la Salud (DeCS), que contiene la traducción del Medical Subject Headings (MeSH) de la Biblioteca Nacional de Medicina (NLM) de Estados Unidos en idiomas español y portugués, ampliado con nuevas categorías de descriptores para permitir una mejor indexación de la literatura científica y técnica de la región, en especial la salud pública y el medio ambiente.

Desde finales de los años 90 y actualmente está en etapas avanzadas de desarrollo, tiene su centralidad en la plena adopción de Internet como los medios de producción y las fuentes de los flujos de información científica y técnica. La cooperación técnica se lleva a cabo a través de la Biblioteca Virtual en Salud (BVS), que fue lanzado en marzo de 1998 con la Declaración de San José Hacia la Biblioteca Virtual en Salud, aprobada durante el 4º Congreso Regional de Ciencias de la Salud (CRICS4). En esta Declaración los representantes de los países reconocen que el acceso a la información constituye uno de los elementos centrales para lograr la equidad en salud; que



las nuevas tecnologías de la información y de la comunicación plantean riesgos y oportunidades para el desarrollo humano en la región y el Centro Latinoamericano y del Caribe en Ciencias de la Salud, bajo el liderazgo de BIREME tiene la capacidad de controlar estas tecnologías, adaptándolos a la realidad de la región. Al mismo tiempo, BIREME cooperar con la Agencia de los Derechos de Protección al Estado de Sao Paulo (FAPESP), Brasil, para la creación de la Biblioteca Científica Electrónica en Línea - SciELO (Scientific Electronic Library On Line), como un modelo para la publicación de cooperación de revistas de calidad en línea en Internet de acceso abierto. (BIREME, 2017)

### **BVS Colombia**

La BVS Colombia como un bien público de información, se enfoca en el resultado del esfuerzo colaborativo de la red de instituciones productoras, intermediarias y usuarias de la información científico técnica en salud del país; con el propósito de apoyar a los entes generadores de información, ofrece la posibilidad de participar en metodologías como LILACS, el más importante y abarcador índice de la literatura científica y técnica en Salud de América Latina y del Caribe, apoyando desde hace 30 años al aumento de la visibilidad, el acceso y la calidad de la información en la Región.

La Fundación Universitaria de Ciencias de la Salud (FUCS) consciente de la importancia de ser parte del modelo BVS, y de apoyar la cooperación entre los actores de la estrategia (productores, intermediarios y usuarios de la información), firmó junto a la Organización Panamericana de la Salud/Organización Mundial de la Salud (OPS/OMS), el «Acuerdo de cooperación técnica para la consolidación y posicionamiento de la BVS de Colombia», este acuerdo designa a la fundación coordinadora del proyecto BVS Colombia, con la responsabilidad y el reconocimiento en el liderazgo de la gestión del sistema de información científica de salud de Colombia, lo que contribuirá de forma esencial a ampliar la visibilidad y mejorar la calidad de la

producción científica y técnica nacional, además de responder al llamado de la Organización Panamericana de la Salud de promoción del uso del conocimiento científico en los procesos de decisión en salud del país. (Niño Téllez, Biblioteca virtual de salud, 2016). En Colombia el modelo BVS es coordinado durante el periodo 2016-2023, por el Sistema de Bibliotecas de la Fundación Universitaria de Ciencias de la Salud- FUCS.

### **Conclusiones**

La vitrina del conocimiento diseñada para la visibilidad y divulgación de información seleccionada por expertos y organizada bajo una metodología estructurada, es una compilación de materiales diseñados por múltiples actores, como organismos de Naciones Unidas, ONG internacionales, entidades de Gobierno y academia, todas ellas basadas en evidencia científica, que permiten facilitar la comprensión de contenidos en áreas relacionadas en salud para apoyar la toma de decisiones soportadas y hacer uso de la tecnología para acercar a los usuarios y romper brechas de acceso a la información.

Con el respaldo de BIREME OPS/OMS, las vitrinas del conocimiento se alojan en las Bibliotecas Virtuales de Salud, portales que tienen como misión la integración de fuentes de información en salud que promueve la democratización y ampliación del acceso a la información científica y técnica en salud en los países de América Latina y el Caribe.

Con el respaldo de BIREME OPS/OMS, las vitrinas del conocimiento se alojan en las Bibliotecas Virtuales de Salud, portales que tienen como misión la integración de fuentes de información en salud que promueve la democratización y ampliación del acceso a la información científica y técnica en salud en los países de América Latina y el Caribe.

## Bibliografía

- CENTRO LATINOAMERICANO Y DEL CARIBE DE INFORMACIÓN EN CIENCIAS DE LA SALUD. (s.f.). BIREME OPS/OMS. Sao Paulo, Brasil: Autor. <[http://www.paho.org/bireme/index.php?option=com\\_content&view=article&id=33:historia&Itemid=215](http://www.paho.org/bireme/index.php?option=com_content&view=article&id=33:historia&Itemid=215)>
- CENTRO LATINOAMERICANO Y DEL CARIBE DE INFORMACIÓN EN CIENCIAS DE LA SALUD. (2020). Guía de la BVS. Sao Paulo, Brasil: Autor. <[https://red.bvsalud.org/modelo-bvs/wpcontent/uploads/sites/3/2020/12/BVS\\_Guia\\_final\\_30set2020\\_tradu%C3%A7%C3%A3o\\_espanhol.pdf](https://red.bvsalud.org/modelo-bvs/wpcontent/uploads/sites/3/2020/12/BVS_Guia_final_30set2020_tradu%C3%A7%C3%A3o_espanhol.pdf)>
- CENTRO LATINOAMERICANO Y DEL CARIBE DE INFORMACIÓN EN CIENCIAS DE LA SALUD. (s.f.). Portal de la red BVS. Sao Paulo, Brasil: Autor. <<https://red.bvsalud.org/es/vitrinas-delconocimiento/>>
- NIÑO TÉLLEZ, A. L. (2016a). Acercamiento a la visibilidad de la información en salud en Colombia y la Biblioteca Virtual de Salud. *Repertorio de Medicina y Cirugía*, 25(4), 228-231. <<https://revistas.fucsalud.edu.co/index.php/repertorio/article/view/104>>
- NIÑO TÉLLEZ, A. L. (2016b). Biblioteca virtual de salud. *Repertorio de Medicina y Cirugía*, 25(4), 201-202. <<https://revistas.fucsalud.edu.co/index.php/repertorio/article/view/106>>

# Avaliação de acessibilidade digital: Biblioteca Digital Brasileira de Teses e Dissertações (BDTD)

## **CÁSSIO TEIXEIRA DE MORAIS**

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA (IBICT)

[cassiomorais@ibict.br](mailto:cassiomorais@ibict.br)

## **DENISE APARECIDA FREITAS DE ANDRADE**

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA (IBICT)

[deniseandrade@ibict.br](mailto:deniseandrade@ibict.br)

## **PRISCILA MACHADO BORGES SENA**

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA (IBICT)

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL (UFRGS)

[priscila.sena@ufrgs.br](mailto:priscila.sena@ufrgs.br)

## **EIXO TEMÁTICO**

Comunicación académica, científica y cultural en abierto

## **RESUMO**

A acessibilidade informacional digital é um tema fundamental na agenda do Instituto Brasileiro de Informação Ciência e Tecnologia (Ibict) para guiar os avanços tecnológicos do Brasil. Neste trabalho se avalia a acessibilidade do Portal da Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), responsável por integrar os sistemas de informação de teses e dissertações existentes nas instituições de ensino e pesquisa do Brasil. Nesse contexto, a instituição promove a democratização dos serviços disponibilizados por meio de boas práticas nos seus ambientes web, de modo que permita a inclusão de pessoas com necessidades especiais no uso da plataforma virtual. Trata-se de uma pesquisa descritiva e experimental com abordagem quantitativa, cujo objetivo é analisar a acessibilidade da BDTD, na qual se recorre à validação automática dos códigos HTML e CSS e aos avaliadores automáticos Access Monitor, ASES

e WAVE. Como resultados se identificaram problemas de acessibilidade presentes em cinco páginas da BDTD: Página inicial, Pesquisa avançada, Sobre (Tede), Rede (como participar) e FAQ. Ressalta-se que, embora a análise automatizada de acessibilidade contribua para visualizar as barreiras existentes, para se adotar os procedimentos de correção e aplicação da acessibilidade é necessário o acompanhamento e análise dos gestores e desenvolvedores do portal.

#### **PALAVRAS-CHAVE**

Acessibilidade digital; avaliadores automáticos; barreiras de acessibilidade; Portal BDTD.

## **1 Introdução**

Ao considerar a acessibilidade digital uma ferramenta vinculada aos elementos provenientes da arquitetura da informação, é possível visualizar um conjunto estrutural que influencia na amplitude informacional por meio da interoperabilidade de sistemas e softwares. A sua correta aplicação beneficia todos os pesquisadores com um ambiente web acessível, em especial possibilita meios para que o conteúdo disponível no Portal da Biblioteca Digital Brasileira de Teses e Dissertações (BDTD) seja reconicionado e interpretado por meio de leitores sonoros de tela ou linguagens de sinais, tornando as informações acessíveis ao público com necessidades especiais visuais e auditivas.

De acordo com Cunha Rocha & Barros David (2020, p. 225), “Um portal é uma página específica na Internet que serve como ponto de acesso direto a outros conjuntos de serviços e informações, contendo subdivisões sobre determinado tema ou área do conhecimento”.

Atualmente a informação acadêmica cresce principalmente por conta dos portais que disponibilizam versões dos trabalhos publicados em formato digital. Esses documentos eletrônicos representam uma etapa da acessibilidade, que promove a divulgação da ciência. Quando falamos em

ciência é preciso raciocinar sobre os processos que envolvem informação e comunicação, ou ainda pensar em um sistema social que “integra elementos que vão desde a figura do pesquisador/cientista/acadêmico ao fluxo de idéias, fatos, teorias, métodos, literatura científica e instrumentos que permitem a operacionalização das investigações” (Targino, 2000, p. 24).

Nessa conjuntura de integração se faz fundamental transpor as dificuldades proporcionando autonomia e independência aos portadores de necessidades especiais por meio da acessibilidade digital, que por sua vez podem ser superadas graças às tecnologias da informação e comunicação. Em contrapartida, mesmo o ambiente web apresentando diversas vantagens afloradas pela tecnologia, ainda sim, manifesta-se que o usuário pode ter dificuldades de acesso, navegação ou não conseguir compreender as informações veiculadas (Rocha, 2013).

A forma como se desenvolve um portal, desde a aparência em design, disposições de textos, imagens e vídeos passa por uma sequência de códigos denominados linguagem de programação, que obedecem a uma estrutura lógica. Essa estrutura fornece caminhos que possibilitam a interoperabilidade com outros sistemas de interação que visam disponibilizar a informação presente no portal em outras formas de compreensão, seja facilitando a visualização ampliando a tela, ou por meio dos já citados softwares leitores de tela, que interpretam por voz ou linguagem de sinais o conteúdo presente em portais como a BDTD.

Uma das plataformas virtuais do Ibict mais conhecidas, a BDTD integra os sistemas de informação de teses e dissertações existentes nas instituições de ensino e pesquisa do Brasil, e também estimula o registro e a publicação de teses e dissertações em meio eletrônico. A BDTD, em parceria com as instituições brasileiras de ensino e pesquisa, possibilita que a comunidade brasileira de C&T publique e difunda suas teses e dissertações produzidas no País e no exterior, dando maior visibilidade à produção científica nacional (BDTD, 2023).

No âmbito nacional, o papel que desempenha o Instituto Brasileiro de Informação Ciência e Tecnologia (Ibict) é fundamental para guiar os avanços tecnológicos no campo informacional do Brasil. Sendo assim é essencial que seus produtos e serviços prestados sejam acessíveis à toda sociedade. Têm-se como objetivo analisar a acessibilidade da BDTD, na qual se recorre à validação automática dos códigos HTML e CSS e aos avaliadores automáticos Access Monitor, ASES e WAVE, que seguem como padrão as normas Web Content Accessibility Guidelines (WCAG) da World Wide Web Consortium (W3C) para avaliar as condições de acessibilidade.. Pressupõe-se que a partir da análise das páginas do Portal da BDTD será possível compreender os problemas de acessibilidade, bem como atuar no sentido de corrigi-los.

## **2 Acessibilidade informacional digital**

A internet é uma das ferramentas que dão suporte à acessibilidade informacional digital, portanto, as ramificações dessa rede envolvem as pessoas em um mundo tecnológico. Segundo Paula & Carvalho (2009, p. 64) “a acessibilidade à informação digital só é possível quando são eliminadas barreiras de comunicação, quando os desenvolvedores de tecnologia digital se preocupam em facilitar o seu acesso a todas as pessoas”. O maior resultado dessa conectividade é a facilidade de interação das pessoas incluídas digitalmente com acesso à internet. Partindo dessa condição existe a possibilidade de ampliar o alcance da informação acadêmica presente em portais da web e contribuir com o progresso da comunicação científica.

A acessibilidade, segundo Sonza et al. (2013), trata-se de um ambiente web livre de barreiras que impossibilite as pessoas fazerem uso de sistemas computacionais. Sendo assim, a acessibilidade informacional digital proporciona acesso ao conteúdo e serviços de forma igualitária para todas as pessoas. Sonza et al. (2013) destacam a internet integrante fundamental da acessibilidade como um dos temas mais estudados e difundidos. O

desenvolvimento proporcionado pelas tecnologias digitais de informação e comunicação favoreceu o aperfeiçoamento do ambiente web, ao eliminar barreiras de acessibilidade e tornar acessível a informação científica aos portadores de necessidades especiais.

### **2.1 Desenvolvimento de ambientes web acessível**

Aliada nesses avanços, a Arquitetura da Informação tem o papel de planejar, organizar e projetar o ambiente web, o que viabiliza a recuperação da informação e a satisfação do usuário, além de contribuir com a gestão e o desenvolvimento de novos conhecimentos (Cunha Rocha, C., & Barros David, P., 2020).

A atividade científica tem como uma de suas principais características a prerrogativa de promover a disseminação dos conhecimentos produzidos por meio da comunicação da informação. Inerente a essa condição, o ambiente web é o canal de comunicação que deve ser avaliado com base em normas estabelecidas, padrões e critérios de acessibilidade regulamentados.

Cunha Rocha & Barros David (2020) afirmam que o processo de avaliação de um ambiente web não é uma tarefa simples, pois está relacionado a atividades de reflexão e planejamento para se atingir objetivos, tendo como base o domínio de conhecimentos e técnicas específicas.

### **2.2 Norma nacional e internacional de acessibilidade**

O compromisso de garantir acessibilidade nos portais da web é um esforço que pode ser mensurado por meios de normas internacionais e nacionais. O interesse é somado ao alcance da internet levando em consideração todas as pessoas que podem ser beneficiadas. A aplicação da acessibilidade nos portais de comunicação científica é um fator de desenvolvimento social.

Em prol desse desenvolvimento para garantir acessibilidade ao ambiente web, um esforço internacional foi realizado pelo World Wide Web Consortium (W3C), comitê internacional formado por grandes empresas da Internet, que atua como gestor de diretivas para a Internet. Dentre as iniciativas do comitê



foi elaborado um guia com diretrizes para acessibilidade do conteúdo web, denominado Web Content Accessibility Guidelines (WCAG) (MPDG et al., 2016, p.4).

Nacionalmente, leis foram implementadas buscando promover a acessibilidade do conteúdo web, principalmente nos portais de entidades governamentais, com o intuito de padronizá-los. Dessa forma, em dezembro de 2005 o Governo Federal, por intermédio do Ministério do Planejamento, Orçamento e Gestão, desenvolveu a cartilha chamada Modelo de Acessibilidade do Governo Eletrônico (eMAG) com a finalidade de ser o norteador no desenvolvimento e adaptação de conteúdos digitais do Governo Federal, garantindo o acesso a todos. É relevante destacar que o eMAG trata de uma versão especializada do documento do WCAG voltado para o governo brasileiro, porém o eMAG não exclui qualquer boa prática de acessibilidade do WCAG (BRASIL, 2014).

### **2.3 Deficiência visual e tecnologia assistiva**

A deficiência visual descrita como cegueira e baixa visão caracteriza-se por uma condição irreversível de perda ou redução da resposta visual em ambos os olhos em caráter permanente, mesmo após tratamento médico, cirurgias ou o uso de lentes corretivas (Instituto Benjamin Constant [IBC], 2005).

Considerando que uma pessoa com necessidades especiais seja resultado de condições atípicas involuntárias, não obstante continuam a viver enfrentando inúmeros obstáculos impostos. A sociedade deve prover formas de assistência e inclusão para que essas pessoas possam ser integradas no ambiente informacional. Atualmente a tecnologia assistiva é uma forma de inclusão, termo que pode ser definido, segundo Fraz et al. (2019, p. 72), como todo produto, equipamento, dispositivo, recurso, estratégia, prática ou serviço que objetive promover a funcionalidade relacionada à atividade e à participação de pessoas com deficiência ou mobilidade reduzida, visando à sua autonomia, independência, qualidade de vida e inclusão social.

A interoperabilidade de sistemas faz parte da funcionalidade da BDTD. Entretanto, os recursos tecnológicos que promovem acessibilidade têm como requisito um segundo passo, a compatibilidade entre softwares com os quais são programados. Tais recursos listados por Nascimento & Silva (2015), são os teclados em Braille e ampliados - facilitam a visualização de pessoas cegas ou com baixa visão, assim como teclados virtuais que são aplicativos ou softwares que atendem a essa função; a lupa eletrônica aumenta textos e imagens auxiliando a compreensão; impressoras em Braille imprimem em papel informações presentes nos sites; leitores de tela captam as informações e disponibiliza em áudio.

### **3 Procedimentos metodológicos**

Este é um estudo descritivo e experimental com abordagem quantitativa, cujo objetivo é analisar a acessibilidade da BDTD. A pesquisa foi dividida em cinco etapas: i) seleção dos avaliadores; ii) seleção das páginas avaliadas; iii) validação do código-fonte; iv) avaliação de acessibilidade digital; e v) tabulação e análise dos dados. Os avaliadores foram selecionados com base nos documentos [Emag \(2014\)](#) e na [Cartilha de acessibilidade na Web do W3C - Brasil](#). Para assegurar maior exatidão aos resultados, escolheu-se um validador HTML, um validador CSS e três softwares avaliadores de acessibilidade digital, a saber: Access Monitor, ASES e WAVE.

<b>Avaliador</b>	<b>Empresa</b>	<b>Diretriz</b>	<b>Resultados</b>	<b>Apresentação</b>
Access Monitor	Governo Federal Brasil	Emag 2014	Erros e alertas	Percentual de acessibilidade e relatório
ASES	Agência para a Modernização Administrativa (AMA)	WCAG 2.1	Aceitáveis, erros e não aceitáveis	Pontuação de acessibilidade e relatório
WAVE	Institute for Disability Research, Policy & Practice	WCAG 2.1	Erros, contraste erros, alertas, características, elementos estruturais e Aria <sup>1</sup>	Indica a contagem dos resultados e apresenta os dados no próprio layout do site

QUADRO 1. Avaliadores automáticos de acessibilidade

Fonte: Elaboração própria, com base nos dados da pesquisa (2023)

Atualmente, o site BDBT oferece acesso aos seus conteúdos por meio de cinco abas principais e agrupa outras informações adicionais em sua página inicial. Nesta pesquisa optou-se por avaliar apenas 5 páginas do site, a saber: 1) Página inicial - coleta os principais dados, estatísticas, informações e serviços oferecidos pelo site; 2) Pesquisa avançada - permite ao usuário limitar sua pesquisa; 3) Sobre (Tede) - descreve o Sistema de Publicação Eletrônica de Teses finais e Dissertativas; 4) Rede (como participar) - descreve o processo de como participar da BDTD; 5) FAQ - informações sobre como entrar em contato com o BDTD.

---

<sup>1</sup>ARIA ou Accessible Rich Internet Applications refere-se a uma especificação do W3C para aprimorar a acessibilidade do site.

As análises ocorreram entre os dias 19 e 25 de abril de 2023, e os relatórios fornecidos pelo software avaliador foram compilados e tabulados. A fim de criar um padrão entre os softwares, a nomenclatura "alerta" foi usada para se referir a todas as informações nomeadas como: "não aceitáveis" e "avisos".

## **4 Resultados e discussões**

Nesta seção é apresentado o relatório de acessibilidade da BDTD que se encontra organizado em dois tópicos: a primeira parte trata do diagnóstico do código HTML e CSS, enquanto a segunda parte aborda a avaliação de acessibilidade por meio de softwares. Em ambos os casos os resultados são exibidos pela quantidade de erros e alertas e a descrição das barreiras identificadas.

### **4.1 Validação da linguagem CSS e HTML**

Lista-se na Tabela 01 a quantidade de erros e avisos de diagnóstico em páginas da BDTD, por meio da validação de código HTML e CSS. Verifica-se que a página com mais erros no HTML foi a "Página inicial" e a com menor número foram as páginas "Sobre - TED" e "Busca avançada". Por sua vez, observa-se que o código CSS não apresenta variações de valores entre as páginas avaliadas, com exceção da página "Busca avançada", que obteve 34 alertas a mais. Estes resultados demonstram que foram utilizadas estruturas semelhantes para o desenvolvimento do código CSS, no entanto existem particularidades na estrutura no código HTML.

	ERRO		ALERTA		TOTAL
	ERRO	ALERTA	ERRO	ALERTA	TOTAL
	18	23	9	252	302
	10	32	9	286	337
	11	36	9	252	308
	11	20	9	252	292
	12	21	9	252	294
<b>Total</b>	<b>62</b>	<b>132</b>	<b>45</b>	<b>1294</b>	<b>1533</b>

TABELA 1. Validação da linguagem de marcação HTML e CSS da BDTD

Fonte: Elaboração própria, com base nos dados da pesquisa (2023)

Nas Tabelas 2 e 3 se visualiza uma perspectiva geral das barreiras diagnosticadas por página nos validadores HTML e CSS. Para compilar os dados na tabela foi necessário agrupar textos similares, omitindo as orientações com o nome do atributo/elemento<sup>2</sup> e função. As áreas em destaque verde referem-se aos resultados que apresentaram um desvio de valores ao comparar as páginas. E os títulos atribuídos ao cabeçalho da tabela como: “Pg. 1” a “Pg.5”, referem-se respectivamente: 1- Página Inicial, 2- Busca Avançada, 3- Sobre: TED, 4- Rede: como participar? e 5- Serviços: FAQ.

---

<sup>2</sup> Um elemento ou tag HTML são hipertextos, que se conectam entre si formando a página. Por sua vez, os atributos HTML são palavras especiais usadas dentro das tags de abertura para controlar o comportamento do elemento. Disponível em: <[https://www.alura.com.br/artigos/o-que-e-html-suas-tags-parte-1-estrutura-basica?utm\\_source=gnarus&utm\\_medium=timeline](https://www.alura.com.br/artigos/o-que-e-html-suas-tags-parte-1-estrutura-basica?utm_source=gnarus&utm_medium=timeline)>

TIPO	BARREIRA DESCRIÇÃO	Pg. 1	Pg. 2	Pg.3	Pg.4	Pg.5
Erro	Uso incorreto de elementos e atributos	4	3	3	3	3
Erro	Valor inválido para o atributo: mídia não reconhecida ou vazia	4	4	4	4	4
Erro	Ausência de texto alternativo para imagens	6	1	1	1	1
Erro	Elementos abertos	3	3	3	3	3
Erro	Elemento obsoleto	1	0	0	0	0
Erro	Marca final de um elemento sem abertura	0	0	0	0	1
Erro	Elemento "Title" vazio	0	0	1	1	1
Alerta	O atributo é desnecessário para recursos JavaScript	14	16	12	12	13
Alerta	O atributo é obsoleto.	1	1	1	1	1
Alerta	O documento não é mapeável para XML 1.0	2	1	1	1	1
Alerta	A barra final não tem efeito e interage mal com valores de atributos	6	14	22	6	6

TABELA 2. Barreiras diagnósticas na validação HTML

Fonte: Elaboração própria, com base nos dados da pesquisa (2023)

De acordo com a Tabela 2, a barreira com maior número de alertas está relacionada com uso de atributos desnecessários para JavaScript e a com falhas na no efeito da barra final. E a barreira com maior número de “erros” refere-se à atribuição incorreta de valores para atributos.

TIPO	DESCRIÇÃO	Pg. 1	Pg. 2	Pg.3	Pg.4	Pg.5
Erro	Presença de muitos valores ou valores não reconhecidos	5	5	5	5	5
Erro	Recurso de mídia obsoleto	1	1	1	1	1
Erro	Escrito de forma incorreta, ou regra proprietária	3	3	3	3	3
Alerta	Propriedade em desuso	12	12	12	12	12
Alerta	Separador inválido	1	1	1	1	1
Alerta	Propriedade proprietária desconhecida	176	210	176	176	176
Alerta	Não é reconhecido como um pseudoelemento/classe proprietário válido	18	18	18	18	18
Alerta	Hack CSS	4	4	4	4	4
Alerta	Regra não está implementada.	3	3	3	3	3
Alerta	Uso de cores em elementos e atributos	38	38	38	38	38

TABELA 3. Barreira diagnósticas na validação CSS por página

Fonte: Elaboração própria, com base nos dados da pesquisa (2023)

Observa-se na Tabela 03 que os aspectos que merecem mais atenção no código CSS são: o uso e a correta aplicação de elementos, propriedades e atributos; a verificação de elementos e atributos descontinuados e a cautela com o uso de cores em elementos e atributos, a fim de não prejudicar a acessibilidade visual do site. Com relação a distribuição, o número de erros e alertas segue constante entre as páginas, com exceção a barreira “propriedade proprietária desconhecida”, que apresentou 34 avisos a mais.

Estas análises permitem identificar que a “página inicial” necessita de mais atenção. Muito embora exista uma variação de erros e alertas, a correção de alguns elementos e atributos tornaria a estrutura do HTML e CSS mais próxima ao padrão W3C.

## 4.2 Avaliação automática de acessibilidade

O Gráfico 1 apresenta o resultado da pontuação de acessibilidade atribuída pelos avaliadores Monitor de Acesso e ASES. Para fins de padronização foi determinado que os resultados serão representados em percentuais. De acordo com o gráfico, há uma diferença de cerca de 20% na pontuação atribuída pelos avaliadores, no qual a BDTD foi, no geral, melhor avaliada pela ASES, indicando que o site está mais alinhado com as recomendações da norma brasileira de acessibilidade (Emag) do que com as diretrizes internacionais, o padrão WCAG. Além disso, cabe destacar que as páginas com desempenho inferior às demais foram: “Sobre – TEDE”, “Rede – Como participar” e “Serviços – FAQ”.

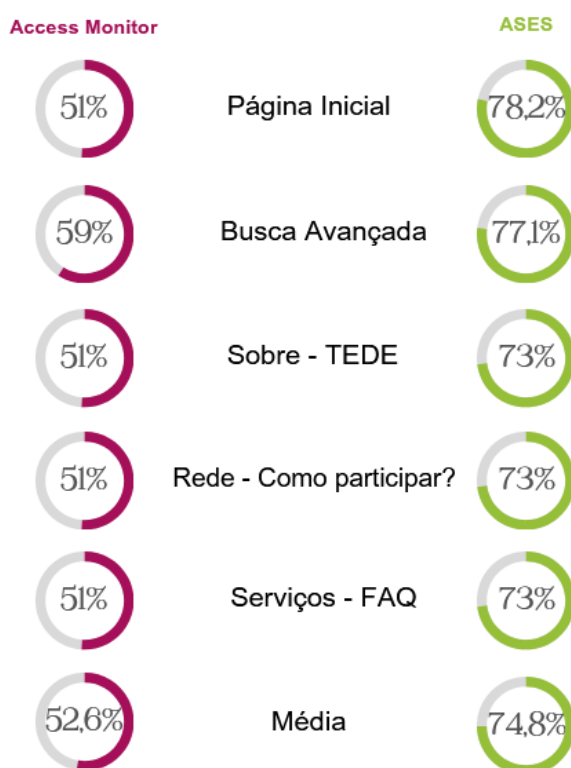


GRÁFICO 1. Pontuação de acessibilidade Access Monitor X ASES

Fonte: Elaboração própria, com base nos dados da pesquisa (2023).



O Gráfico 2 apresenta a comparação da quantidade de erros e alertas identificados por página por cada avaliador. Os resultados em azul referem-se ao Access Monitor, enquanto os de rosa ao ASES e os verdes ao WAVE. Ao todo foram encontrados 656 erros e alertas, no entanto este valor pode representar problemas repetidos, uma vez que contém a informação dos 3 avaliadores. Constatou-se que, embora a “Página inicial” e “Busca Avançada” apresentem uma das pontuações de acessibilidade mais altas no Gráfico 1, no Gráfico 2, elas são as páginas com maior número de erros e alertas. Além disso, o avaliador ASES foi o que mais diagnosticou problemas, sendo 98 erros e 317 alertas.

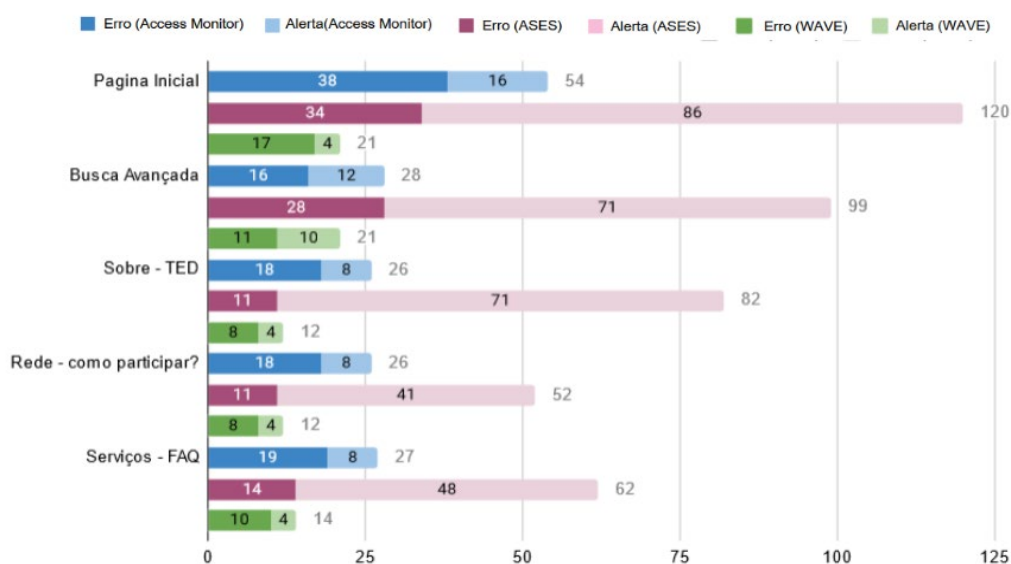


GRÁFICO 2. Quantidade de alertas e erros por páginas

Fonte: Elaboração própria, com base nos dados da pesquisa (2023).

O Quadro 2 exibe o relatório por avaliador das recomendações de acessibilidade que não foram constatadas no site da BTDT. Para compilar os dados utilizou como base a Diretriz - WCAG 2.1, uma vez que o Emag também possui relação com a diretriz internacional, sendo possível correlacionar os dados.

<b>Cr�terios</b>	<b>Descri�o</b>	<b>Ac cess</b>	<b>A SES</b>	<b>WA VE</b>
<b>1.1.1</b>	Alternativas de texto para conte�do n�o textual	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>1.2.1, 1.2.2, 1.2.3</b>	Alternativas para m�dias baseadas em tempo			<input type="checkbox"/>
<b>1.3.1</b>	Organiza�o o c�digo HTML e CSS	<input type="checkbox"/>	<input type="checkbox"/>	
<b>1.4.3</b>	Contraste m�nimo	<input type="checkbox"/>		<input type="checkbox"/>
<b>1.4.4</b>	Redimensionamento do texto	<input type="checkbox"/>		
<b>2.1.2 e 2.1.1</b>	Oper�vel atrav�s de teclado		<input type="checkbox"/>	
<b>2.4.1</b>	Mecanismo para ignorar blocos	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>2.4.2</b>	T�tulo descritivo e informativo		<input type="checkbox"/>	<input type="checkbox"/>
<b>2.4.4</b>	Link possui uma finalidade em contexto	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>2.4.6</b>	Cabe�alhos e r�tulos para formul�rio e outros elementos			<input type="checkbox"/>
<b>2.4.9</b>	Descreve links clara e sucintamente	<input type="checkbox"/>	<input type="checkbox"/>	
<b>2.4.10</b>	Utiliza cabe�alhos da sess�o	<input type="checkbox"/>	<input type="checkbox"/>	
<b>3.1.1</b>	Utiliza idioma da p�gina	<input type="checkbox"/>		
<b>3.2.2, 3.2.3 e 3.2.5</b>	Navega�o consistente e sem altera�o de contexto	<input type="checkbox"/>	<input type="checkbox"/>	
<b>3.3.2</b>	Utiliza r�tulos ou instru�es para orientar a entrada de dados	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>4.1.1 e 4.1.2</b>	Respeitar os Padr�es Web	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

QUADRO 2. Crit rios WCAG identificados para an lise por avaliador

Fonte: Elabora o pr pria, com base nos dados da pesquisa (2023)

Conforme o Quadro 2, as barreiras identificadas referem-se a: I) apresenta o de conte do, tais como: aus ncia de alternativas de texto para imagens, a incorreta identifica o e descri o de links, a aus ncia ou uso de r tulos pouco inform ticos; II) design do site, tais como: baixo contraste,

problemas como redirecionamento do texto, navegação inconsistente; III) desenvolvimento do código, tais como: organização do código e correta aplicação dos padrões Web.

## 5 Considerações finais

Conclui-se que o objetivo geral da pesquisa foi atingido ao identificar problemas presentes no Portal da BDTD. Esta análise é importante para ter a dimensão das barreiras de acesso existentes para os usuários, as quais prejudicam especialmente aqueles indivíduos com deficiência visual. Realizou-se a análise de cinco páginas da BDTD por meio de três avaliadores de acessibilidade digital. As incompatibilidades constatadas foram apresentadas e comparadas em tabelas e gráficos. Por sua vez, os resultados demonstram que ainda existem muitas barreiras a serem superadas para atingir níveis mais altos de acessibilidade.

Algumas das ações que podem contribuir para melhorar a qualidade do site são: se atentar as normas estabelecidas para utilização dos códigos HTML e CSS, manter um código organizado e atualizado, utilizar textos alternativos para descrever imagens, utilizar uma linguagem clara e simples para descrever os links, usar cores com contraste que facilitem a visualização dos conteúdos, utilizar rótulos e textos que orientem os usuários para o uso de ferramentas e entradas de dados.

Após todo o processo de análise, verifica-se que as páginas da BDTD apresentam repetidas ocorrências com os mesmos tipos de erros e avisos, por isso é essencial que o modelo de layout padrão do portal seja bem elaborado e siga as recomendações internacionais e nacionais de acessibilidade. Neste sentido, a análise automatizada de acessibilidade foi importante para o diagnóstico de falhas e limitações do portal. Conseqüentemente, o desenvolvimento de um portal dentro dos padrões de acessibilidade colabora com a disseminação e comunicação da informação científica digital, uma vez

que a BDTD disponibiliza o acesso à Teses e Dissertações de várias instituições do Brasil.

Por fim, é importante ressaltar que embora a análise automatizada de acessibilidade contribua para visualizar as barreiras existentes, para se adotar os procedimentos de correção e aplicação da acessibilidade é necessário o acompanhamento e análise dos gestores e desenvolvedores do portal. Ademais, é necessário que após as correções os gestores estabeleçam uma rotina de manutenção e análise de acessibilidade digital.

## Bibliografia

- BRASIL. (2014). Departamento de Governo Eletrônico. eMAG: Modelo de Acessibilidade em Governo Eletrônico. Versão 3.1. <<http://emag.governoeletronico.gov.br/>>
- BIBLIOTECA DIGITAL BRASILEIRA DE TESES E DISSERTAÇÕES [BDTD] (2023). Sobre a BDTD. Disponível em: <<https://bddd.ibict.br/vufind/>>
- CUNHA ROCHA, C., & BARROS DAVID, P. (2020). Avaliação da arquitetura da informação em portais de periódicos: uma discussão teórica. *Perspectivas Em Gestão & Conhecimento*, 10(2), 223-243. <<https://periodicos.ufpb.br/index.php/pgc/article/view/50239>>
- FRAZ, J. N., HOTT, D. F. M., MOREIRA, G. E., & RODRIGUES, G. M. (2019). Tecnologia Assistiva: produtos e serviços disponíveis na internet. *Ponto de Acesso*, 13(3), 70–84. DOI: <<https://doi.org/10.9771/rpa.v13i3.35225>>
- INSTITUTO BENJAMIN CONSTANT [IBC]. (2005). Definindo a cegueira e a visão subnormal. <<http://www.ibc.gov.br/?itemid=396>>
- MINISTÉRIO DO PLANEJAMENTO, DESENVOLVIMENTO E GESTÃO [MPDG], SECRETARIA DE TECNOLOGIA DA INFORMAÇÃO [STI], & DEPARTAMENTO DE GOVERNO DIGITAL [DGD]. (2016). Boas Práticas para Acessibilidade Digital na Contratação de Desenvolvimento WEB - versão 1.0. <<https://emag.governoeletronico.gov.br/cartilha-contratacao/#melhorando>>
- NASCIMENTO, M. O. DO, & SILVA, E. F. DA. (2015). E-acessibilidade em bibliotecas: uma análise sobre disponibilidade, direito e limitações do acesso à informação na web. *Bibliocanto*, 1(1), 45-67. <<https://Brapci.inf.br/index.php/res/v/120247>>

- PAULA, SONIA NASCIMENTO DE; CARVALHO, JOSÉ OSCAR FONTANINI DE. (2009). Acessibilidade à informação: proposta de uma disciplina para cursos de graduação na área de biblioteconomia. *Ciência da Informação*, 38(3), 64-79. DOI: <https://doi.org/10.1590/S0100-19652009000300005>
- ROCHA, JANICY APARECIDA PEREIRA. (2013). (In)acessibilidade na web para pessoa com deficiência visual: um estudo de usuários à luz da cognição situada. 2013. 157 f. Dissertação (Mestrado em Ciência da Informação) –Programa de Pós-Graduação em Ciência da Informação, da Escola de Ciência da Informação, da Universidade Federal de Minas Gerais, Belo Horizonte, 2013. [https://repositorio.ufmg.br/bitstream/1843/ECIC-9BFKDS/1/disserta\\_\\_o\\_janicy\\_rocha\\_2013.pdf](https://repositorio.ufmg.br/bitstream/1843/ECIC-9BFKDS/1/disserta__o_janicy_rocha_2013.pdf)
- SONZA, ANDRÉA POLETTO; KADE, ADROVANE; FAÇANHA, AGEBSON; REZENDE, ANDRÉ LUIZ ANDRADE, NASCIMENTO, GLEISON SAMUEL DO; ROSITO, MAURÍCIO COVOLAN; BORTOLINI, SIRLEI; FERNANDES, WOQUITON LIMA (orgs.). (2013). Acessibilidade e tecnologia assistiva : pensando a inclusão sociodigital de PNEs. Bento Gonçalves, RS: Instituto Federal do Rio Grande do Sul Campus Bento Gonçalves, 2013. cap. 5, p. 313-364. (Série Novos Autores da Educação Profissional e Tecnológica). <https://cta.ifrs.edu.br/livro-acessibilidade-e-tecnologia-assistiva-pensando-a-inclusao-sociodigital-de-pessoas-com-necessidades-especiais/>
- TARGINO, M. D. G. (2000). Comunicação científica: uma revisão de seus elementos básicos. *Informação & Sociedade: Estudos*, 10(2). <http://hdl.handle.net/20.500.11959/brapci/92195>

# Criação de Metadados Dublin Core para documentos de patentes no Arca – Repositório Institucional da Fundação Oswaldo Cruz

## **ÉDER DE ALMEIDA FREYRE**

FUNDAÇÃO OSWALDO CRUZ

INSTITUTO DE COMUNICAÇÃO E INFORMAÇÃO CIENTÍFICA E TECNOLÓGICA EM SAÚDE

[eder.freyre@iciict.fiocruz.br](mailto:eder.freyre@iciict.fiocruz.br)

## **FABRICIA PIRES PIMENTA RIBEIRO**

FUNDAÇÃO OSWALDO CRUZ

INSTITUTO CARLOS CHAGAS

[fabricia.pimenta@fiocruz.br](mailto:fabricia.pimenta@fiocruz.br)

## **CLAUDETE FERNANDES DE QUEIROZ**

FUNDAÇÃO OSWALDO CRUZ

INSTITUTO DE COMUNICAÇÃO E INFORMAÇÃO CIENTÍFICA E TECNOLÓGICA EM SAÚDE

[claudete.queiroz@iciict.fiocruz.br](mailto:claudete.queiroz@iciict.fiocruz.br)

## **RAPHAEL BELCHIOR RODRIGUES**

FUNDAÇÃO OSWALDO CRUZ

INSTITUTO DE COMUNICAÇÃO E INFORMAÇÃO CIENTÍFICA E TECNOLÓGICA EM SAÚDE

[raphael.belchior@iciict.fiocruz.br](mailto:raphael.belchior@iciict.fiocruz.br)

## **EIXO TEMÁTICO**

Comunicación académica, científica y cultural en abierto

## RESUMO

O trabalho apresenta como foi a construção da coleção de Patentes para o Arca - Repositório Institucional da Fundação Oswaldo Cruz - Fiocruz. Os documentos de patente apresentam características peculiares, que justificam o desenvolvimento e a validação de um novo conjunto de metadados específico. O processo envolveu a definição do padrão de metadados, campos e critérios relevantes sobre essa tipologia entendendo que as patentes são documentos considerados cruciais para a inovação e competitividade da Instituição e em termos gerais, contém uma descrição técnica da invenção, apresentando o estado da arte da tecnologia, o problema técnico a ser suplantado e a solução alcançada pela invenção. Dentro deste contexto, a Fiocruz, a partir da década de 80 criou o Núcleo de Gestão Tecnológica (GESTEC) para zelar pelo patrimônio científico e tecnológico em relação a proteção das patentes. As patentes são uma importante fonte de informação para a pesquisa e o desenvolvimento tecnológico, pois apresentam informações detalhadas sobre os processos inventivos. A criação dos metadados de patentes permitiu uma melhor gestão dos dados já que essas informações podem ser usadas por pesquisadores e empresas para identificar patentes relevantes, entender a natureza da invenção protegida e avaliar a sua relevância para a sociedade e a Instituição a qual pertence. É nesse sentido que este trabalho explica por que esta coleção foi criada para depósito no Repositório Arca.

## PALAVRAS-CHAVE

Patentes; Arca – Repositório institucional da Fiocruz; Gestão do conhecimento.

## Introdução

A Fundação Oswaldo Cruz vinculada ao Ministério da Saúde no Brasil, com 123 anos de criação, tem em suas premissas a promoção da saúde, o desenvolvimento social, a geração e a disseminação do conhecimento científico e tecnológico. Neste contexto, cumprindo sua missão, a fundação se destaca no atendimento à população brasileira por meio do Sistema Único de Saúde (SUS). Além de produzir insumos, kits diagnósticos e vacinas e oferecer serviços de assistência ambulatorial especializada e hospitalar, ainda possui 32 programas de pós-graduação stricto sensu em áreas de Ciências da Saúde e Biológicas, uma escola de nível técnico e vários programas lato sensu, e uma

grande atuação na produção de pesquisa científica e desenvolvimento tecnológico na área da Saúde.

Toda produção intelectual da Fiocruz é estimulada, conforme descrito na **Política de Acesso Aberto ao Conhecimento**, a ser depositada no Repositório Institucional Arca, visando reunir, hospedar, disponibilizar e dar visibilidade à produção intelectual. O Repositório é organizado por comunidades e coleções que abrigam o conhecimento científico produzido das Unidades Técnico-Científicas da Fiocruz (Fundação Oswaldo Cruz, 2023).

Dentre as diversas tipologias documentais produzidas na Fiocruz, tem-se o documento de patente. Em termos gerais, o documento de patente contém uma descrição técnica da invenção, apresentando o estado da arte da tecnologia, o problema técnico a ser suplantado e a solução alcançada pela invenção. Nesse sentido, o documento de patente detém grande quantidade de pesquisa e conhecimento técnico-científico, na medida em que disponibiliza a informação mais recente sobre o estado da arte e, ainda, oferece informações de caráter legal e comercial. O conceito de Patente segundo o Instituto Nacional de Propriedade Industrial - INPI é um título de propriedade temporário, oficial, concedido pelo Estado, por força de lei, ao seu titular ou seus sucessores (pessoa física ou pessoa jurídica), que passam a possuir os direitos exclusivos sobre o bem, seja de um produto, de um processo de fabricação ou aperfeiçoamento de produtos e processos já existentes, objetos de sua patente. Terceiros podem explorar a patente somente com permissão do titular (mediante uma licença) (Instituto Nacional...2023).

A Fiocruz a partir da década de 80, com a observação da importância acerca da proteção da propriedade intelectual, cria um núcleo de gestão tecnológica<sup>1</sup>

---

<sup>1</sup> Atualmente o Núcleo de Gestão Tecnológica passou a se chamar Coordenação de Gestão Tecnológica.



para zelar pelo patrimônio científico e tecnológico em relação a concessão das patentes. “Entre as áreas tecnológicas nas quais a Fiocruz possui depósito de pedidos de patente, destacam-se as de vacinas e medicamentos. Outras áreas são kits para diagnósticos, equipamentos e dispositivos médico-hospitalares, bioinseticidas e outros” (Macedo; Barbosa, 2000). Atualmente, segundo dados do **Observatório da Fiocruz**, a Fundação detém 216 tecnologias protegidas via sistema de patentes, com 729 documentos de patente. Destaca-se ainda que a Fiocruz apresenta o perfil de proteção de suas tecnologias não somente no Brasil, mas também em outros países ou regiões, tais como Estados Unidos da América, Europa, Índia, China, Rússia, Austrália, Peru, Colômbia, Chile e Canadá.

Dentro deste contexto, é notório que tal conjunto de documentos de patente seja uma importante fonte de informação para a pesquisa e o desenvolvimento tecnológico, pois proporciona informações detalhadas sobre processos pelos quais os inventores alcançaram a solução para o problema técnico, apresentando a experimentação laboratorial que consubstancia a solução proposta. Além disso, a descrição no documento de patente deve ser feita de forma clara e suficiente pelos inventores, de modo a possibilitar sua realização por técnico no assunto e, ainda, indicar sua melhor forma de execução (Pimenta, 2017).

E, exatamente pelo fato de conter grande quantidade de informação, que por vezes não esteja disponível em outro meio de divulgação, tanto por ser publicada exclusivamente no sistema de patentes, quanto por estar na fronteira do conhecimento (Pimenta, 2017).

É nesse sentido que o presente trabalho contextualiza a criação da coleção de documentos de patente, de titularidade da Fiocruz, em seu repositório institucional, e apresenta metadados específicos para esta tipologia documental.

## Metodologia

Em 2013, a equipe do [Núcleo de Inovação Tecnológica do ICICT](#) solicitou à Coordenação Executiva do Repositório Arca a criação da Coleção de Patentes para o depósito deste tipo de documento. Dada às peculiaridades desta tipologia documental, os metadados classicamente utilizados para artigos científicos não contemplavam importantes aspectos dos documentos de patente, tais como, a indicação de titularidade; as diferentes versões e numerações do documento ao longo do processo de tramitação processual; as datas de depósito, publicação e concessão; a indicação da classificação internacional; a indicação do documento prioritário; a indicação dos países de depósito e concessão do documento de patente.

Em seguida, iniciou-se o estudo dos metadados adequados para representar as idiosincrasias desta tipologia com o objetivo de fornecer informações em um formato estruturado e consistente para facilitar a descrição, a recuperação e a preservação dos documentos no Repositório Arca, respeitando os diferentes tipos de documentos de patente, versões, publicações e prazos (tanto para publicação quanto para expiração) (Souza; Vendrusculo; Melo, 2000).

A partir desta etapa, iniciou-se uma análise comparativa de metadados de diferentes repositórios para esta tipologia. Inicialmente, utilizou-se o [Repositório Digital de Publicações Científicas da Universidade de Évora](#) (Portugal) como referência. Em seguida, utilizou-se o [Web Ranking of Repositories](#) para busca e análise comparativa com outros repositórios. Ainda foram realizadas comparações com repositórios reconhecidamente importantes de outras instituições brasileiras de ciência e tecnologia, com vocações de pesquisa semelhantes às da Fiocruz. As buscas foram realizadas em 22/03/2023.

## Resultados e Discussão

A criação de repositórios institucionais pode ser um desafio posto que deve refletir a natureza e peculiaridades da instituição. Dentre os diferentes tipos documentais que refletem a produção intelectual da instituição, os documentos de patente devem refletir o conhecimento científico e desenvolvimento tecnológico a ser protegido para uma eventual futura exploração comercial (Pimenta, 2017). Nesse sentido, dada a natureza e propósito peculiares desta tipologia, a Coordenação Executiva do Repositório Arca criou a Coleção de Patentes com metadados específicos. Os campos, para descrição desta tipologia documental baseados no padrão Dublin Core (Tabela 1), permitiram que as informações coletadas pudessem definir e permitir a busca de diferentes versões e traduções dos documentos de patente. Há ainda outras contribuições que se destacam na recuperação da informação quando se trata de documentos de patente, tais como as datas e diferentes identificações alfanuméricas, na linha do tempo de tramitação, quer seja no Brasil ou em outros países, de um documento de patente.

O estudo desses metadados específicos ainda permitiu a categorização de cada campo como obrigatório ou não obrigatório e repetitivo ou não repetitivo, o que torna tal conjunto de metadados adequado tanto para documentos de patente prioritários quanto para famílias de patentes com pedidos de diferentes origens, ou mesmo se ainda em fase de pedido ou já concedido ou com proteção expirada. Ainda permitiu-se identificar em quais países a tecnologia estava em processo de proteção e as datas de concessão e expiração de tal proteção em cada país.

<b>IDENTIFICAÇÃO</b>	<b>METADADOS</b>	<b>CAMPO</b>		<b>NOTAS</b>
Título	dc.title	Obrigatório	Não Repetitivo	Preencha as informações solicitadas.
Inventor	dc.contributor.inventor	Obrigatório	Repetitivo	Entre com os nomes dos inventores do Item.
E-mail	dc.creator.email	Não Obrigatório	Repetitivo	Entre com o(s) respectivo(s) e-mail(s) do(s) inventor(es).
Afiliação	dc.creator.affiliation	Obrigatório	Repetitivo	Descreva corretamente a afiliação do(s) autor(es).
Instituição Titular	dc.identifier.sharewith	Obrigatório	Repetitivo	Entre com a respectiva instituição titular.
Palavras-Chave	dc.subject.other	Obrigatório	Repetitivo	Entre com as palavras-chave.
Tipo	dc.type	Obrigatório	Não Repetitivo	Selecione o(s) tipo(s) de conteúdo.
Número de depósito	dc.identifier.ndeposito	Obrigatório	Não Repetitivo	Entre com o número de depósito.
Data de depósito	dc.date.deposito	Obrigatório	Não Repetitivo	Entre com a data de depósito.
Número de depósito da prioridade	dc.identifier.ndepositoprior	Não Obrigatório	Repetitivo	Entre com o número de depósito da prioridade.
Número da publicação	dc.identifier.npublicacao	Obrigatório	Não Repetitivo	Entre com o número da publicação.
Data do	dc.date.issued	Obrigatório	Não	Informe somente o ano que consta

<b>IDENTIFICAÇÃO</b>	<b>METADADOS</b>	<b>CAMPO</b>		<b>NOTAS</b>
Documento			Repetitivo	no documento.
Número da carta patente	dc.identifier.ncartapatente	Não Obrigatório	Não Repetitivo	Entre com o número da carta patente.
Data de concessão	dc.date.concessao	Não Obrigatório	Não Repetitivo	Entre com a data de concessão da patente.
Data de expiração	dc.date.expiracao	Não Obrigatório	Não Repetitivo	Entre com a data de expiração da patente.
Identificador	dc.identifier	Obrigatório	Repetitivo	Entre com o identificador.
Referência	dc.identifier.citation	Obrigatório	Não Repetitivo	Inclua a referência bibliográfica, segundo NBR 6023.
Resumo	dc.description.abstract	Não Obrigatório	Repetitivo	Entre com o(s) resumo(s) do Item.
Idioma	dc.language.iso	Obrigatório	Não Repetitivo	Selecione o idioma do conteúdo principal do Item.
Direito Autoral	dc.rights	Obrigatório	Não Repetitivo	Preencha os direitos deste Item como open access, restricted access ou closed access
Países de depósito da patente	dc.description.deposito	Obrigatório	Repetitivo	Entre com as notas pertinentes do país de depósito.
Países de concessão da patente	dc.description.concessao	Não Obrigatório	Repetitivo	Entre com as notas pertinentes de concessão da patente.

IDENTIFICAÇÃO	METADADOS	CAMPO		NOTAS
Financiador	dc.description.sponsorship	Não Obrigatório	Repetitivo	Entre com o financiador, patrocinador, agência de fomento ou nome da bolsa.
Notas	dc.description	Não Obrigatório	Repetitivo	Entre com outra descrição ou comentários neste campo.

TABELA 1. Descrição dos metadados de documentos de patente criados para o [Repositório Arca](#)

No Repositório Arca, são 10 comunidades da Fiocruz que possuem a coleção Patentes: Centro de Desenvolvimento Tecnológico em Saúde, Instituto Gonçalo Moniz, Instituto Rene Rachou, Instituto Carlos Chagas, Instituto Aggeu Magalhães, Instituto de Tecnologia em Fármacos, Instituto de Tecnologia em Imunobiológicos, Instituto Nacional de Infectologia Evandro Chagas, Instituto Nacional de Saúde da Mulher, da Criança e do Adolescente Fernandes Figueira e Instituto Oswaldo Cruz, com um total de 59 documentos disponibilizados em acesso aberto. Na Figura 1, apresentamos um exemplo de patente depositada no Repositório.

**ANTICORPO, SEU USO, COMPOSIÇÃO FARMACÉUTICA O COMPREENDENDO, MÉTODO DE DIAGNÓSTICO DE INFECÇÕES FÚNGICAS, KIT DE DIAGNÓSTICO DE INFECÇÕES FÚNGICAS E MÉTODO PARA TRATAMENTO DE INFECÇÕES FÚNGICAS**  
 Figueiredo, Alexandre Bezerra Conde et al. | Data do documento: 2021

**Título alternativo**  
 Antibody, related use, pharmaceutical composition including method for diagnosing fungal infections, fungal infection diagnosis kit and method for treating fungal infections

**Inventor(es)**  
 Figueiredo, Alexandre Bezerra Conde  
 Rodrigues, Marcio Lourenço  
 Conte, Fernando de Paiva  
 Fonseca, Fernanda Lopes  
 Antisawa, Marcia

**Afiliação**  
 Fundação Oswaldo Cruz. Instituto de Tecnologia em Imunobiológicos. Rio de Janeiro, RJ, Brasil  
 Fundação Oswaldo Cruz. Instituto Carlos Chagas. Curitiba, PR, Brasil. Fundação Oswaldo Cruz. Centro de Desenvolvimento Tecnológico em Saúde. Rio de Janeiro, RJ, Brasil.  
 Fundação Oswaldo Cruz. Instituto de Tecnologia em Imunobiológicos. Rio de Janeiro, RJ, Brasil.  
 Fundação Oswaldo Cruz. Centro de Desenvolvimento Tecnológico em Saúde. Rio de Janeiro, RJ, Brasil.  
 Fundação Oswaldo Cruz. Instituto de Tecnologia em Imunobiológicos. Rio de Janeiro, RJ, Brasil.

**Instituição titular da patente**  
 Fundação Oswaldo Cruz

**Número de depósito**  
 PCT/BR2021/05007

**Data de depósito**  
 2021-12-12

**Número de depósito da prioridade**  
 BR10202002185-6A

**Número de publicação**  
 WO2021/151180A2

**Resumo**  
 A presente invenção fornece anticorpos monoclonais contra oligômero de quitina através da técnica de hibridomas. Os referidos anticorpos podem ser utilizados como ferramentas para diagnóstico e tratamento de infecções fúngicas. Também são reveladas composições farmacêuticas e kits para tratamento de infecções fúngicas compreendendo os referidos anticorpos. Ademais são revelados ainda um método de diagnóstico de infecções fúngicas utilizando os referidos anticorpos e o seu uso na preparação de um medicamento para tratar infecções fúngicas.

**Resumo em inglês**  
 The present invention provides monoclonal antibodies against chitin oligomer using the hybridoma technique. Said antibodies can be used as tools for diagnosing and treating fungal infections. Pharmaceutical compositions and kits for treating fungal infections including said antibodies are also disclosed. A method for diagnosing fungal infections using said antibodies and the use thereof in the preparation of a drug for treating fungal infections are also disclosed.

**Palavras-chave**  
 Cryptococcus  
 Anticorpo

**Palavras-chave em inglês**  
 Antibody  
 Cryptococcus

**Editor**  
 Organização Mundial de Propriedade Intelectual

**Número do IPC**  
 A61K

**País de depósito da patente**  
 Brasil  
 Organização Mundial para Propriedade Intelectual OMPI

**Referência**  
 FIGUEIREDO, Alexandre Bezerra Conde et al. Anticorpo, seu uso, composição farmacêutica o compreendendo, método de diagnóstico de infecções fúngicas, kit de diagnóstico de infecções fúngicas e método para tratamento de infecções fúngicas. WO Pat. 2021/151180 A2; 05 ago. 2021. 97p.

FIGURA 1

<<https://www.arca.fiocruz.br/handle/icict/48830>>

A partir do conjunto de metadados específico, pode-se compará-lo com outras iniciativas semelhantes. A coleta no **Web Ranking of Repositories** e a troca de experiência com instituições conhecidas no escopo brasileiro permitiu observar alguns repositórios com coleções específicas para patentes. Alguns exemplos de iniciativas podem ser explorados no Brasil e no exterior, a partir da Tabela 2.

O número de documentos de patente depositados nesses repositórios destaca que o conhecimento gerado e publicado nas patentes, depositado e preservado nos repositórios pode ser um diferencial para a instituição, estimulando a geração de novos conhecimentos e inovações.

REPOSITÓRIOS INSTITUCIONAIS	LINK DE ACESSO	NÚMERO DE OBJETOS	QUANTIDADE PATENTES DEPOSITADAS	PAÍS	INSTITUIÇÃO
Pantheon	<a href="https://pantheon.ufrj.br/handle/11422/129">https://pantheon.ufrj.br/handle/11422/129</a>	19.705	188	Brasil	UFRJ
Repositório Institucional UNESP	<a href="https://repositorio.unesp.br/handle/11449/108949">https://repositorio.unesp.br/handle/11449/108949</a>	201.673	470	Brasil	UNESP
ATTENA Repositório Digital da UFPE	<a href="https://repositorio.ufpe.br/handle/123456789/34552">https://repositorio.ufpe.br/handle/123456789/34552</a>	48.511	119	Brasil	UFPE
Repositório Institucional UFVJM	<a href="http://acervo.ufvjm.edu.br">acervo.ufvjm.edu.br</a>	2.654	20	Brasil	UFVJM
LUME	<a href="https://lume.ufrgs.br/">https://lume.ufrgs.br/</a>	246.588	491	Brasil	UFRGS
Instituto Federal de Santa Catarina	<a href="http://repositorio.ifsc.usp.br/">http://repositorio.ifsc.usp.br/</a>	2.260	105	Brasil	IFSC
Shocker Open Access Repository – SOAR	<a href="https://soar.wichita.edu/">https://soar.wichita.edu/</a>	18.669	45	EUA	Wichita State University
Rice University's digital scholarship archive	<a href="https://scholarship.rice.edu/">https://scholarship.rice.edu/</a>	77.258	582	EUA	Rice University
HKU Scholars Hub	<a href="https://hub.hku.hk/advanced-search?location=patent">https://hub.hku.hk/advanced-search?location=patent</a>	230.883	1116	Hong Kong	University of Hong Kong
<a href="http://ruj.uj.edu.pl/">Jagiellonian University Repository</a>	<a href="http://ruj.uj.edu.pl/">ruj.uj.edu.pl/</a>	238.533	27	Polônia	Jagiellonian University
Universitas Airlangga Repository	<a href="https://repository.unair.ac.id">https://repository.unair.ac.id</a>	103.836	170	Indonésia	Airlangga University
<a href="https://re.public.polimi.it/">RE.PUBLIC@POLIMI</a>	<a href="https://re.public.polimi.it/">https://re.public.polimi.it/</a>	152.029	1073	Itália	<a href="http://www.politecnico.mi.it/">Politecnico di Milano</a>



REPOSITÓRIOS INSTITUCIONAIS	LINK DE ACESSO	NÚMERO DE OBJETOS	QUANTIDADE PATENTES DEPOSITADAS	PAÍS	INSTITUIÇÃO
VTechWorks	<a href="https://vtechworks.lib.vt.edu/">https://vtechworks.lib.vt.edu/</a>	95.552	125	EUA	<a href="#">Virginia Tech</a>
University of Muhammadiyah Malang Institutional Repository	<a href="https://eprints.umm.ac.id">https://eprints.umm.ac.id</a>	105.340	191	Indonésia	University of Muhammadiyah Malang

TABELA 2. Instituições com coleções específicas para documentos de patente em repositórios institucionais

### Considerações finais

O presente trabalho, a partir do entendimento de que os documentos de patente apresentam características peculiares, quis apresentar como são as características da tipologia Patentes e o motivo de criação dos metadados específicos que melhor descrevem tal coleção. Inicialmente, a busca por iniciativas semelhantes identificou que poucos repositórios institucionais desenvolveram esta coleção para documentos de patente, demonstrando a negligência para com a gestão do conhecimento deste tipo de conhecimento.

A criação de metadados permitiu uma melhor gestão dos dados, pois o usuário pode em um único acesso obter dados específicos, tais como, datas de depósito, publicação e concessão; indicação da classificação internacional; indicação do documento prioritário; indicação dos países de depósito e concessão do documento de patente. Essas informações podem ser usadas por pesquisadores e empresas para identificar patentes relevantes, entender a natureza da invenção protegida e avaliar a sua relevância para a sociedade e a Instituição a qual pertence, como também preservar e disseminar este relevante conteúdo informacional.

## Bibliografía

- FREYRE, ÉDER DE ALMEIDA. (2019). Manual de tratamento de dados: preenchimento de metadados para entrada no Arca - Repositório Institucional da Fiocruz. 3. ed. rev. e aum. Rio de Janeiro: FIOCRUZ/ICICT, 2019. [422] p. il. <<https://www.arca.fiocruz.br/handle/icict/13257>>
- FUNDAÇÃO OSWALDO CRUZ. (2014). Política de Acesso Aberto ao Conhecimento. Portaria da Presidência n. 329/2014. <<https://portal.fiocruz.br/documento/politica-de-acesso-aberto-ao-conhecimento-da-fiocruz>>
- FUNDAÇÃO OSWALDO CRUZ. (2023). Patentes. Rio de Janeiro, 2023. <<https://portal.fiocruz.br/patentes>>
- INSTITUTO NACIONAL DA PROPRIEDADE INTELECTUAL. (2023). Patentes. Rio de Janeiro, 2023. <<https://www.gov.br/inpi/pt-br/servicos/patentes>>
- MACEDO, M. F. G.; BARBOSA, A. L. F. (2000). Patentes, pesquisa & desenvolvimento: um manual de propriedade intelectual. Rio de Janeiro: Editora Fiocruz, 2000. <<https://books.scielo.org/id/6tmww>>
- PIMENTA, F. P. (2017). A patente como fonte de informação (des)necessária para a Biotecnologia em Saúde. *Transinformação*, 29(3), 319-328. <<https://www.arca.fiocruz.br/handle/icict/27061>>
- SOUZA, M. I. F.; VENDRUSCULO, L. G.; MELO, G. C. (2000). Metadados para a descrição de recursos de informação eletrônica: utilização do padrão Dublin Core. *Ciência da Informação*, 29(1), 93-102.

# Dados abertos para a sustentabilidade: avaliação usando os princípios FAIR

## **CRISTINA OLIVEIRA ROMANIN**

UNIVERSIDADE FEDERAL DE SÃO CARLOS (UFSCAR)

[cromanin@estudante.ufscar.br](mailto:cromanin@estudante.ufscar.br)

## **ARIADNE CHLOE MARY FURNIVAL**

UNIVERSIDADE FEDERAL DE SÃO CARLOS (UFSCAR)

[chloe@ufscar.br](mailto:chloe@ufscar.br)

### **EIXO TEMÁTICO**

Dados Abertos

### **RESUMO**

Os dados ambientais são a base para projetos, políticas e iniciativas em prol da sustentabilidade. Surge, então, o questionamento se esses dados estão disponíveis para uso legítimo por organizações, ONGs, e pela comunidade ambientalista. Diante do colocado, a pesquisa propõe identificar e avaliar o grau de encontrabilidade, acessibilidade, interoperabilidade e reusabilidade – ou seja, aplicar os critérios “FAIR”, usualmente aplicados exclusivamente a dados de pesquisa científica – a conjuntos de dados abertos porque podem contribuir ao desenvolvimento sustentável e às metas da Agenda 2030 da ONU. Para isso, foi realizada uma pesquisa documental nos sites de cada Órgão Federal Ambiental e no Portal Brasileiro de Dados Abertos. Foram identificados diversos conjuntos de dados ambientais, entre eles 166 que são objeto de avaliação da pesquisa. Para a análise, foi elaborado um protocolo na forma de checklist adaptado tanto dos princípios FAIR quanto do documento Boas Práticas para Dados na Web da W3C. Por fim, a ferramenta FUJ-I foi utilizada para comparar uma análise manual a uma análise automática, ambas baseadas nos princípios FAIR. Os resultados parciais apontam que a maioria dos dados ambientais do governo brasileiro poderiam ter otimizações em seu grau de abertura e reusabilidade.

**PALAVRAS CHAVE**

Princípios FAIR; Dados Abertos; Dados Abertos Governamentais; Dados Ambientais. FAIR principles; Open Data; Open Government Data; Environmental Data.

# Competência em Dados de Pesquisa: analisando as práticas dos profissionais da informação envolvidos no apoio à pesquisa

## **ANA JULIA LOPES**

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL (UFRGS)

[lopesanajuliacl@gmail.com](mailto:lopesanajuliacl@gmail.com)

## **CATERINA GROPOSO PAVÃO**

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL (UFRGS)

[caterina@cpd.ufrgs.br](mailto:caterina@cpd.ufrgs.br)

### **EIXO TEMÁTICO**

Datos abiertos

### **RESUMO**

A premissa de que os dados de pesquisa são valiosos e requerem tempo e recursos para serem criados, preservados e gerenciados, exige dos pesquisadores habilidades, contudo, nem todos reconhecem todas as etapas da gestão de dados de suas pesquisas. À vista disso, surge a Competência em Dados de Pesquisa como uma atividade dedicada à compreensão das habilidades, conhecimentos e rotinas relativas aos dados de pesquisa e constitui-se, entre outras coisas, em atividades práticas de manipulação de dados de pesquisa em diferentes disciplinas. A proposta apresenta um estudo inicial da dissertação em desenvolvimento no Programa de Pós-Graduação em Ciência da Informação, da Universidade Federal do Rio Grande do Sul (UFRGS), e tem por objetivo analisar as práticas de capacitação em dados de pesquisa promovidas pelas universidades internacionais. Para obter os resultados iniciais utilizou-se o Webometrics.info 2022 para seleção do corpus de análise, e a pesquisa documental como método para identificar os trabalhos desenvolvidos pelas universidades selecionadas.

## **PALAVRAS CHAVE**

Competência em dados; capacitação em dados; dados de pesquisa. Data literacy; data training; research data.

## **Introdução**

Os dados são um dos principais tópicos e tendências na pesquisa da informação. O paradigma intensivo de dados da ciência e da pesquisa tem sido explicado como uma tendência emergente que influencia o trabalho de informação. As inúmeras ferramentas, como repositórios, portais de periódicos, diretórios e outras redes de compartilhamentos, contribuem para o crescimento da produção científica e uso de dados (Biazon & Marin, 2016).

Com a extensa quantidade de dados, o processo desde o planejamento e organização, até a disponibilização para o reuso deve ser realizado por um profissional que tenha competência para trabalhar com grandes volumes de dados. Dessa forma, surgiu, nos últimos anos, o “cientista de dados”, ou ainda, o bibliotecário de dados. Estes profissionais estariam capacitados, entre outras atividades, a minerar massas desordenadas de dados e produzir conhecimento por serem capazes de discernir o valor de grandes volumes de dados, e gerenciá-los de acordo com suas capacidades e necessidades.

Com esse crescimento exponencial dos dados surgiram diversos desafios relativos à gestão ao longo de todo o ciclo de vida dos dados. Um dos problemas apontados na literatura diz respeito à necessidade de um processo de gerenciamento dos dados no que tange a sua visualização e aproveitamento, considerado um pré-requisito para o seu compartilhamento eficaz entre uma comunidade científica específica e comunidades que não fazem parte do grupo alvo. Também, os cientistas de todas as áreas do conhecimento começaram a ter dificuldades para gerenciar e armazenar o enorme volume de dados gerados em suas próprias pesquisas (Monteiro & Sant’ana, 2018; Lima, Pinto & Farias, 2020).

A respeito disso, Whitmire, Boock, Sutton (2015) afirmam que as complexidades em formatos de dados, padrões, infraestrutura, etc., representam desafios para os pesquisadores atenderem aos requisitos de gerenciamento de dados de pesquisa. A facilidade com que coletam complexos conjuntos de dados está superando seus conhecimentos e habilidades para gerenciá-los adequadamente, habilidades essas, que são cruciais para garantir a qualidade, integridade, capacidade de compartilhamento, descoberta e reutilização dos dados ao longo do tempo de forma econômica, segura e eficiente. Assim sendo, um plano de gestão de dados (PGD) que oriente os pesquisadores durante a coleta e tratamento até a divulgação dos conjuntos de dados é imprescindível para que sejam cuidados ao longo do seu ciclo de vida.

Outro ponto importante, é um sistema confiável para a publicação de dados baseada em uma potente infraestrutura digital viabilizando o compartilhamento de dados de forma eficaz para que esses recursos atinjam todo seu potencial de reuso. Nesse sentido, os Princípios FAIR se apresentam como requisitos orientadores, uma vez que a adoção desses princípios pela comunidade científica permite a incorporação de boas práticas para a publicação compartilhamento e gestão de dados de pesquisa (Henning et al., 2019a).

No entanto, nem todos os pesquisadores reconhecem todas as etapas da gestão de dados de suas pesquisas, e outros não possuem habilidades com ferramentas e softwares, por isso, torna-se fundamental que o pesquisador e demais profissionais que trabalham com dados tenha domínio de todas as etapas da gestão dos dados de pesquisa, isto é, seja versado em utilizar os dados, não somente para saber coletá-los, mas gerenciá-los e disponibilizá-los para os mais diversos fins.

À vista disso, surge a Competências em Dados, que está relacionado à compreensão que o indivíduo tem para lidar com os dados e o que se refere às suas tipologias. Este termo segue a tendência de ver a competência em dados como uma prática sociocultural com o uso de tecnologias digitais no ambiente

da informação. No contexto dos dados de pesquisa, as competências em dados e habilidades são imprescindíveis para o bibliotecário e pesquisador se conscientizar da importância de estarem sempre se atualizando e ampliando suas competências no desenvolvimento de suas atividades, possibilitando assim, um aprendizado a longo prazo para sua vida. É nisso que se constitui a competência em dados, estudo contínuo sobre os dados e seus processos.

Sendo assim, o presente estudo aborda uma análise preliminar das práticas adotadas para desenvolver competências necessárias afim de trabalhar com dados de pesquisa em universidades internacionais, principalmente aquelas que dispõem de centro de capacitação em dados.

### **Dados de pesquisa**

Com o Big Data surgiram estudos sobre a compreensão e uso de dados, um dos principais tópicos e tendências na pesquisa da informação. Seu paradigma intensivo tem sido explicado como um movimento emergente que influencia o trabalho de informação. Os dados podem ser gerados em diversos ambientes e por diversos motivos. Podem ganhar novas atribuições durante o seu ciclo de vida. Desta forma, questiona-se o que é um dado e quando ele pode ser considerado um dado de pesquisa?

Os dados sempre foram presentes e constantes nas pesquisas, principalmente de cunho científico. Com a expansão do uso das tecnologias e possibilidades de buscar mecanismos mais elaborados para realizar pesquisas, e a capacidade de armazenar grandes quantidades de dados em redes de infraestrutura digital, despertou nos pesquisadores das ciências naturais e sociais, das artes e humanidades o interesse nos dados de pesquisa (Koltay, 2017; Reis, 2019).

Com a consolidação da ciência aberta os dados de pesquisa começam a ser reconhecidos como elementos importantes, principalmente a partir do momento em que documentos acadêmicos se tornam objetos digitais



complexos, onde novos serviços e ferramentas são desenvolvidos para produzir, publicar, disseminar e gerenciar esses conceitos de mídia acadêmica, dando a entender que os padrões e fluxos da comunicação científica estão rapidamente se ajustando às novas normas de publicação e compartilhamento de dados de pesquisa (Sayão & Sales, 2020; Silveira, 2021).

O entendimento sobre o termo “dato” pode variar consideravelmente entre pesquisadores e áreas do conhecimento. Semeler e Pinto (2019) escrevem que a palavra “dato” significa uma única peça de informação. Para os autores podem ser fatos ou estatísticas reunidas para referência ou análise de algo; também podem ter uma coleção diversa incluindo imagens, fluxo de vídeos ou áudio, software, algoritmos, etc. O relatório da Organização para a Cooperação e Desenvolvimento Econômico (OCDE) publicado em 2007, que define os dados de pesquisa como registros factuais como pontuações numéricas, registros textuais, imagens e sons, usados como fontes primárias para pesquisa científica e que são normalmente aceitos na comunidade científica como necessários para validar os resultados da pesquisa.

Corroborando com o ponto de vista dos autores supracitados, Sales e Sayão (2019) propõe que os dados de pesquisa são todo e qualquer tipo de registro apurado, observado, ou usado pela pesquisa científica, tratado ou aceito como necessário para validar os resultados pela comunidade científica. No entanto, esse termo não compreende os cadernos de laboratório, análises preliminares e rascunhos de artigos científicos, planos para pesquisas futuras, revisões por pares ou comunicações pessoais com colegas ou objetos físicos.

Os dados de pesquisa têm sido considerados um dos principais produtos das pesquisas acadêmicas nas universidades e centros de pesquisas. Contudo, faz-se necessário iniciativas práticas, principalmente quanto ao uso desses dados para o desenvolvimento de novas pesquisas. Pesquisadores enfrentam novos desafios no gerenciamento de informações de tal maneira que as bibliotecas e seus profissionais começam a se especializar a fim de

promoverem pluralidade de serviços de suporte em dados, tais como treinamentos, instruções, gerenciamento, curadoria e visualização. (Reis, 2019).

Rice e Southall (2016) corroboram que os bibliotecários e pesquisadores, em seus espaços de trabalho, estão aderindo a essas perspectivas, expandindo e iniciando uma série de atividades de apoio à pesquisa, fortalecendo e conscientizando-se sobre questões de dados dentro das instituições e seus benefícios de gerenciar ativamente os dados de pesquisa.

Com base nisso, a competência em dados de pesquisa ajuda que esses profissionais acessem, interpretem, avaliem criticamente, gerenciem, manipulem e usem os dados de forma ética (Prado & Marzal, 2013) compreendendo as nuances da pesquisa baseada em dados.

### **Competência em dados de pesquisa**

Os dados de pesquisa geralmente são integrados às infraestruturas de dados e tornam-se componentes do ambiente de informações. Para o uso adequado de infraestruturas de dados e informações em disciplinas de pesquisa, é necessário conhecimento contínuo de como usá-los. Nesse contexto, pode-se determinar a competência em dados de pesquisa como o conhecimento de uma série de atividades de processamento de dados, desde a identificação até o planejamento, armazenamento, análises, avaliação, gerenciamento, provisão e proteção (Steinerová & Ondrisová, 2019).

Os tipos de dados de pesquisa podem incluir big data, dados abertos, dados qualitativos e quantitativos, dados experimentais, observacionais, sensoriais, simulações, referenciais, entre outros, dos mais variados tipos e formatos de acordo com as especificidades das áreas do conhecimento. A competência em dados ajuda na resistência contra os efeitos nocivos de más práticas intencionais e não intencionais. Definir indesejabilidade é uma questão cultural ou política; contudo, reconhecer a urgência de desenvolver competência em

dados e encontrar consenso sobre questões críticas sobre dados é crucial para navegar em um período de transformação.

As práticas de dados de pesquisa são atividades de manipulação de diferentes disciplinas. Este termo segue a tendência de ver a competência em dados como uma prática sociocultural com o uso de tecnologias digitais no ambiente da informação. No contexto deste estudo, o termo “competência” compreende, então, o sentido de “saber fazer” que requer um conjunto de “habilidades” e implica um posicionamento diante daquilo que se apresenta como necessário. Isto é, ser competente em dados tem o mesmo significado que ser competente em informação, pois é necessária às pessoas de todas as áreas, de todos os níveis educacionais [...] já que tem por objetivo propiciar a capacitação contínua dos indivíduos, que, ao buscarem os conteúdos informacionais [...] tornam-se mais autônomos (Vitorino, 2008; Mata, 2014, p. 16).

Dessa forma, a Competência em dados faz parte no processo de gestão de dados de pesquisa, pois os bibliotecários e pesquisadores precisam acompanhar os estudos e efeitos dos dados nas comunidades científicas, levando-se em consideração a formação de novas habilidades como condensar dados, combiná-los com outras fontes de informação do conhecimento, determinar quando os dados são necessários, como manipular, contextualizar, dentre outros processos (Costa, Sales & Zattar, 2020).

Em outros termos, a competência em dados de pesquisa é uma capacitação para pesquisadores e bibliotecários ultrapassar o desafio de identificar os métodos e técnicas fundamentais para o gerenciamento e análise de grandes volumes de dados.

Com essa perspectiva, surge a necessidade do profissional da informação para auxiliar no processo de desenvolvimento de habilidades para que os pesquisadores tenham mais controle sobre os dados produzidos e saibam, entre outras coisas, documentar todos os procedimentos envolvidos no ciclo

de vida dos dados, possibilitando trabalhar com os dados de forma consciente em qualquer ambiente organizacional. Esse profissional é o Bibliotecário de Dados.

Bibliotecário de dados é um termo novo e sua aplicação encontra-se nos princípios de práticas tradicionais do bibliotecário aos atuais recursos de dados, e emergiu conforme demandavam as tecnologias digitais e a internet na década de 1990 (REIS, 2019). Os bibliotecários estão cada vez mais envolvidos com uma ampla gama de iniciativas relacionadas a dados, incluindo avaliações para determinar as necessidades de instrução relacionadas a dados. Conforme Deja et al. (2021) eles podem usar habilidades e competências em dados para construir programas de aprendizagem para integrar com o corpo docente, alunos e pesquisadores, e demonstrar o fruto de seus trabalhos, para evitar uma situação em que os resultados de suas pesquisas na academia sejam imperceptíveis.

Em pesquisa de Burness, Mann e Neville (2020) sobre a competência em dados por meio de uma comunidade de aprendizado de bibliotecários, os autores destacam que o envolvimento do bibliotecário de dados com as comunidades de aprendizagem oferece oportunidades para: aumentar a compreensão das perspectivas pedagógicas do corpo docente, experiência e necessidades de instrução e programas de bibliotecários; desenvolver relacionamentos com professores e funcionários com reuniões regulares na biblioteca e outros espaços do campus que possam levar a uma variedade de oportunidades colaborativas de ensino e aprendizagem; integrar bibliotecários como parceiros do corpo docente como co-desenvolvedores de currículo em áreas centrais, como alfabetização de informações e dados; e, finalmente, influenciar o ensino e os resultados dos alunos no nível institucional e profissional.

Percebe-se que um papel importante é desempenhado pelos bibliotecários de dados que implementam princípios de competência informacional para gerenciamento de dados no contexto acadêmico e científico. A competência

em dados de pesquisa visa desenvolver o uso de melhores práticas no que se refere aos dados de pesquisa para benefício do próprio pesquisador, do grupo de pesquisa e da instituição à qual pertence, e da comunidade científica como um todo, uma vez que com a variedade de dados e objetos de informação, é fundamental o indivíduo ter independência, saber ler, entender e interpretar os dados para reuso ou reprodutibilidade das pesquisas.

### **Metodologia**

Esta pesquisa trata-se de um estudo de natureza básica, com abordagem qualitativa, de caráter descritivo e exploratório, utilizando-se, principalmente, da pesquisa documental como procedimento metodológico. As etapas da análise documental, nos websites das universidades internacionais, foram: pré-análise (baseada no objetivo da pesquisa); organização (seleção e classificação dos documentos conforme as categorias pré-definidas) e análise (interpretação dos dados e conclusão).

Utilizou-se o ranking do Webometrics.info 2022 para selecionar o corpus de análise constituído por universidades internacionais que dispõe de centro de capacitação de dados, para este piloto selecionaram-se as três primeiras universidades citadas no ranking, com o intuito de analisar as práticas em competência em dados. O estudo deu-se em dois momentos repetido para cada uma das universidades selecionadas:

- 1) verificou-se se as universidades dispunham de centro/setor/divisão de capacitação para competência em dados de pesquisa;
- 2) coletaram-se informações disponíveis no site: documentos, manuais, treinamentos, apresentações, folders de divulgação ou qualquer outro instrumento utilizado como estratégias para a competência em dados. Todo o material relevante para a pesquisa foi salvo em pastas do Google Drive identificadas pelo nome da universidade e

- 3) analisou-se o material coletado e classificaram-se as informações nas categorias previamente organizadas em uma planilha Excel.

As categorias construídas foram: Quem é o público alvo das atividades de desenvolvimento de competências em dados? Quais os tipos de materiais e ferramentas que utilizam para desenvolver as competências em dados? Como ocorrem as práticas para desenvolver as competências em dados (cursos, oficinas)? Quem promove as estratégias para desenvolver as competências em dados? Quais as competências desenvolvidas? A coleta dos dados ocorreu entre os dias 24 e 25 de janeiro de 2023. No Quadro 1 estão as universidades selecionadas, na ordem que se apresentavam no ranking, país e o site.

Nº	PAÍS	UNIVERSIDADE	WEBSITE
1	EUA	Universidade de Harvard	<a href="https://www.harvard.edu/">https://www.harvard.edu/</a>
2	EUA	Universidade de Stanford	<a href="https://www.stanford.edu/">https://www.stanford.edu/</a>
3	UK	Universidade de Oxford	<a href="https://www.ox.ac.uk/">https://www.ox.ac.uk/</a>

QUADRO 1. Universidades selecionadas para análise

### Resultados iniciais

#### *Análise em relação ao público alvo das atividades de desenvolvimento de competências em dados*

Buscando conhecer para quem as universidades direcionam suas capacitações em dados, identificou-se que é dirigida para pesquisadores, funcionários, alunos da própria instituição e comunidade não acadêmica. No entanto, como cada instituição tem suas políticas de extensão, há diferenças do seu público alvo. A Harvard University direciona suas capacitações para os pesquisadores de sua instituição, fazendo parceria com membros da rede de profissionais de dados de Harvard, especialistas no assunto (mesmo que não sejam membros da instituição) e provedores de serviços e recursos a fim de

oferecer um suporte em dados satisfatórios. A Stanford University volta suas atividades de capacitação para seus alunos, pesquisadores e qualquer pessoa interessada nos aspectos de gerenciamento e visualização de dados. Membros do corpo docente da instituição também são contemplados com cursos caso tenham interesse em disponibilizar seus dados para suas publicações. Já a Oxford University dirige suas atividades de treinamento em dados principalmente para alunos, funcionários da Universidade, e pais e educadores de crianças de 6 a 10 anos, bem como para as próprias crianças. Essa capacitação voltada para os pais e crianças é desenvolvida pelo projeto KOALA, sendo de fundamental importância, uma vez que ajuda os pais e as crianças a se prepararem para um mundo cada vez mais movido por dados, e o fato de ter habilidades em ler, entender, transformar, comunicar e integrar dados são tão cruciais quanto a capacidade tradicional de ler e escrever, afirma Batista (2022).

### ***Análise dos tipos de materiais e ferramentas que utilizam para desenvolver as competências em dados***

Visto que as Universidades abrem suas capacitações para alunos, docentes, pesquisadores e pessoas de fora da comunidade científica que tenham interesse em aprender sobre dados (como a Stanford e Oxford com o projeto KOALA), procurou-se identificar quais os tipos de materiais e ferramentas que elas utilizam. Para esse reconhecimento levou-se em consideração os materiais de apoio tais como slides, lista de exercícios, apostilas, tutoriais ou algum outro utilizado pela instituição. Em relação às ferramentas que manuseiam para oferecer as capacitações, considerou-se aquelas online como Youtube, Zoom, ou outro próprio da entidade, e websites, a saber:

A Harvard University utiliza materiais virtuais como por exemplo slides pelo fato de ser mais prático para apresentar o conteúdo proposto. A maioria das capacitações são realizadas virtualmente via Zoom, com algumas opções presenciais. Para aqueles que não conseguem assistir no dia, são disponibilizadas gravações pelo Youtube e no site do departamento/unidade

que oferece a oficina ou curso. Na Stanford University também utilizam materiais virtuais e contam com a plataforma do Youtube para deixar em acesso aberto as oficinas oferecidas. Além dos Slides, a instituição disponibiliza no website da biblioteca e departamento que atua na capacitação de dados, links com arquivos de cada Módulo contendo o conteúdo abordado na sessão. Por fim, na Oxford University é comum ter slides de apresentação, apostilas, folha de tarefas, gravações e tutoriais em vídeo disponibilizados no Youtube.

Pôde-se constatar que das três universidades todos utilizam ferramentas virtuais para oferecer as capacitações, como o Youtube, e slides como material de apoio. Vale ressaltar que durante a análise percebeu-se que, com o retorno das atividades presenciais após o isolamento decorrente da COVID-19, as instituições continuaram com as atividades online e abriram para mais vagas conforme o aumento da procura, seja por parte dos membros das universidades ou comunidade externa. Por meio disso nota-se que ao implementar capacitações sobre dados, as universidades como produtora de conhecimento, assegura ao público subsídios para trabalhar com dados independente da área de atuação, declara Silveira (2021).

### ***Análise de como ocorrem as práticas para desenvolver as competências em dados***

Como viu-se na análise anterior, as ferramentas virtuais são as mais utilizadas para oferecer as capacitações, enquanto que os materiais de apoio predominante são os slides. No entanto, faz-se necessário conhecer de que maneira as universidades oferecem seus serviços.

Se tratando da Harvard University as capacitações são oferecidas por meio de serviços de consulta de gerenciamento de dados, aulas, workshops, seminários sobre uma variedade de tópicos relacionados ao gerenciamento de dados de pesquisa, e webinars com explicações sobre dados de pesquisa. A Stanford University oferece treinamentos em workshops, consultas individuais,



além de vídeos disponibilizados no Youtube pela biblioteca da universidade de Stanford. A Oxford University atua com palestras, workshops, serviços e recursos de treinamento em fontes de informação, consultoria online para gerenciamento de dados de pesquisa e cursos via Zoom.

Nesta seção de análise de como se dá as capacitações nas universidades percebe-se que todos oferecem cursos, oficinas e workshops para tratar de assuntos sobre os dados. A promoção de capacitações em dados pelas instituições vai de encontro com a afirmação de Haendel, Vasilevsky e Wirzet (2012) que declaram que, criar uma cultura onde pesquisadores abordam dados de pesquisa de forma adequada e eficiente contribui na sua formação científica com educação em alfabetização de dados. Shorish (2015) ainda corrobora dizendo que essas habilidades trabalhadas nas oficinas de alfabetização de dados são relevantes mesmo que os alunos não continuem seus estudos para obter graus mais avançados.

### ***Análise de quem promove as estratégias para desenvolver as competências em dados***

A Harvard University conta com a “rede de Bibliotecas Harvard” que oferece serviços de dados para suas comunidades de pesquisa locais, isto é, fornece aos departamentos e bibliotecas de Harvard serviços de consulta de dados de pesquisa, curadoria, instrução e suporte a projetos ajudando a garantir que os dados de pesquisa multidisciplinares sejam localizáveis, acessíveis, interoperáveis e reutilizáveis (FAIR). Dentro dessa rede, tem bibliotecas parceiras que fornecem serviços de dados especializados para suas comunidades de pesquisa locais. Um exemplo a ser citado é a Biblioteca Countway, uma das bibliotecas acadêmicas de pesquisa em saúde mais abrangentes do país, que em resposta às necessidades não atendidas sobre gerenciamento de dados de pesquisa biomédico e biológico oferece aulas e workshops sobre uma variedade de tópicos relacionados ao gerenciamento de dados de pesquisa, como visto na seção anterior.

Na Stanford University quem promove as capacitações também são as bibliotecas da universidade. Assim como em Harvard, cada biblioteca especializada ministra capacitações de serviços de dados e gerenciamento individualmente, conforme a necessidade de seus alunos e profissionais, sejam eles pós-graduandos ou professores. Dentro do site da biblioteca da instituição tem os serviços de apoio à pesquisa, e neles tem suporte para software e serviços para ciência de dados, Serviços de gerenciamento de dados, Serviços de dados de pesquisa. Aqui destaca-se o Centro de Soluções Oceânicas que trabalha com pesquisadores de Stanford em colaboração com outras universidades a fim de avançar na compreensão dos desafios dos dados oceânicos.

Já a Oxford University conta com a “The Compliance Team” (Equipe de Cumprimento) que é responsável por monitorar a adequação dos arranjos da Universidade para conformidade legislativa e apoiar a política de Proteção de Dados da Universidade, incluindo orientações sobre tratamento e proteção de dados. Outro departamento que promove as capacitações em Oxford, é o grupo interdepartamental “Research Data Oxford” (Dados De Pesquisa Oxford) que fornece suporte para todos os aspectos de gerenciamento de dados de pesquisa, desenvolvimento e fornecimento de infraestrutura e serviços de RDM. A Biblioteca Bodleian, que é a principal biblioteca de pesquisa da Universidade de Oxford, também desenvolve atividades, além do projeto KOALA. Interessante destacar esse projeto, pois diferente de todas as instituições, essa tem um trabalho voltado aos pesquisadores de saúde mental, pais e educadores de crianças pequenas, bem como com as próprias crianças, para entender melhor o impacto da coleta de dados pessoais por aplicativos móveis sobre o bem-estar e a saúde delas.

O fato dos tablets, smartphones, computadores e até mesmo videogame estar se tornando o principal meio para as crianças acessarem a Internet, poucos estudos examinam como crianças menores de 11 anos percebem e lidam com a privacidade de dados pessoais durante suas interações com

essas tecnologias móveis. Esse projeto da Universidade de Oxford contribui justamente para preparar os pais e as crianças a trabalhar e reconhecer os dados como parte do dia a dia deles.

Diante de todas essas análises, é imprescindível verificar quais os conteúdos abordados e trabalhos nessas capacitações.

### ***Análise das competências desenvolvidas***

Na Harvard University os assuntos mais trabalhados nas capacitações são referentes aos dados FAIR e como torná-los localizáveis, acessíveis, interoperáveis e reutilizáveis; como selecionar um repositório para seus dados; como e onde publicar seus dados e como realizar um gerenciamento de dados de pesquisa com eficiência. Os principais assuntos abordados na Stanford University, são sobre práticas recomendadas de dados - incluindo gerenciamento de recursos de dados, limpeza de conjuntos de dados, normalização e padronização de conjuntos de dados e bancos de dados, visualização de dados - incluindo princípios básicos na comunicação geral de dados e gestão de Dados. Por fim, a Oxford University discute em suas capacitações maneiras de aproveitar os dados dentro dos governos, como melhorar os sistemas de dados e as oportunidades para pesquisas relevantes para políticas. Também debatem sobre proteção de dados e como mantê-los seguros no ambiente de trabalho, a cerca de manipulação e aquisição de dados bem como seus compartilhamentos (que engloba toda seleção e preparação de dados para compartilhar citando os princípios FAIR para torná-los encontráveis, acessíveis, interoperáveis e reutilizáveis) e workshops que abrangem fontes de dados para pesquisa - descoberta, acesso e uso.

Mediante o exposto, percebe-se que as Universidades tem como foco comum de suas capacitações o gerenciamento de dados e como os princípios FAIR auxiliam nesse processo. Não é sem motivo que esses dois assuntos são predominantes nos planos de ensino das instituições.

A gestão apropriada de dados é um pré-requisito para um compartilhamento de dados eficaz em uma comunidade científica, afirma Monteiro e Sant'Ana (2018). Para os autores, as instituições científicas e acadêmicas passaram a ter mais responsabilidades no gerenciamento de dados científicos, uma vez que no âmbito dos serviços de apoio à pesquisa, o gerenciamento dos dados é uma preocupação que cabe à biblioteca e departamentos de ensino participar mais ativamente deste contexto. Isto é, os serviços de apoio à pesquisa, aos dados produzidos, e a interação entre os bibliotecários, especialistas no assunto e professores contribui para que os programas de capacitação resultem em habilidades para as comunidades. (Sant'anna, Dias & Maculan, 2019).

Nesse contexto de gerenciamento de dados, os princípios FAIR se apresentam como requisitos orientadores, que “[...] devem ser aceitos pela comunidade de produtores e consumidores de dados de pesquisa, com a finalidade de integrar boas práticas para a publicação e compartilhamento de dados científicos”. (Henning et al., 2019, p. 398). No entanto, sabe-se que incorporar os princípios FAIR no processo de gerenciamento de dados não é algo tão simples, principalmente ao que diz respeito a interoperabilidade pois solicita desenvolvimento de estruturas de apoio como ontologias e tesouros, “[...] padronizando conteúdo de campos de metadados com vocabulários controlados[...] ” e muitas instituições não têm esse amparo tecnológico (Shintaku; Appel & Oliveira, 2021, p. 145).

Compreende-se com essas afirmações que as capacitações desenvolvidas pelas universidades são um pontapé inicial para que as pessoas aprimorem suas habilidades e se tornem competentes em dados.

## Considerações finais

A aquisição de habilidades de competência em dados é um problema para pesquisadores, estudantes de pós-graduação, e comunidade não acadêmica que precisam se tornar trabalhadores científicos competentes no que diz respeito aos dados. Tendo em vista os aspectos apresentados neste estudo, considera-se que o objetivo de analisar as práticas de capacitação em dados de pesquisa promovidas pelas universidades de âmbito internacional foi atingido.

Conferindo os resultados deste estudo piloto é impossível não comparar com a realidade brasileira. Infelizmente as universidades do país ainda não dão o devido destaque à competência em dados, sejam de pesquisa, governamentais ou de áreas específicas. Porém, há iniciativas como a do GT RDP-Brasil, que têm elaborado diversos materiais relacionados às práticas de dados de pesquisa e que podem ser utilizados pela comunidade de pesquisadores e bibliotecários. Na análise de literatura realizada para embasar este estudo foram localizadas publicações sobre gerenciamento de dados e competência em dados de pesquisa (Monteiro & Sant'anna, 2018; Sousa et al., 2010; Presser & Silva, 2018; Lima et al., 2020) que abordam o tema de forma teórica, mas não como é a prática nas universidades. Os resultados preliminares demonstram que as Universidades internacionais estão empenhadas em capacitar suas comunidades, uma vez que as tecnologias digitais e o constante volume de dados produzidos requer que o indivíduo se capacite para trabalhar com esses ativos de informação.

## Bibliografia

- BATISTA, A. F. M. (2022). Alfabetização em dados é competência essencial para qualquer profissão. <https://www.insper.edu.br/noticias/alfabetizacao-em-dados-e-competencia-essencial-para-qualquer-profissao/>

- BIAZON, T. & MARIN, T. (2016, 22 Julho). Ciência aberta: uma nova forma de fazer ciência. DiCYT Plataforma Iberoamericana de Divulgação. <<https://www.dicyt.com/noticia/ciencia-aberta-uma-nova-forma-de-fazer-ciencia>>
- BURNES, T., MANN, E., & NEVILLE, T. (2020). Exploring data literacy via a librarian-faculty learning community: a case study. *The Journal of Academic Librarianship*, 45. <<https://www.sciencedirect.com/science/article/pii/S0099133319304069?via%3Dihub#s0090>>
- COSTAL, M., SALES, L., & ZATTAR, M. (2020). Competência em dados: habilidades na atuação e formação do bibliotecário. *Biblos*, 34(2). DOI: <<https://doi.org/10.14295/biblos.v34i2.11809>>
- DEJA, M. ET AL. (2021). The Impact of Basic Data Literacy Skills on Work-Related Empowerment: The Alumni Perspective. *College & Research Libraries*, 82(5), 708-729. <<https://crl.acrl.org/index.php/crl/article/view/25016/32878>>
- HAENDEL, M. A.; VASILEVSKY, N. A., & WIRZET, J. A. (2012). Dealing with Data: a Case Study on Information and Data Management Literacy. *PLoS Biology*, 10(5). <<https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001339>>
- HENNING, P. C., ET AL. (2019a). Desmistificando os princípios FAIR: conceitos, métricas, tecnologias e aplicações inseridas no ecossistema dos dados FAIR. *Pesq. Bras. em Ci. da Inf. e Bib.*, 14(3), 175-192. <<https://brapci.inf.br/index.php/res/v/150613>>
- HENNING, P. C. ET AL. (2019b). GO FAIR e os princípios FAIR: o que representam para a expansão dos dados de pesquisa no âmbito da Ciência Aberta. *Em Questão*, 25(2), 389-412. <<https://seer.ufrgs.br/EmQuestao/article/view/84753>>
- KOLTAY, T. (2017). Data literacy for researchers and data librarians. *Journal of Librarianship and Information Science*, 49(1), 3-14. <<https://journals.sagepub.com/doi/abs/10.1177/0961000615616450>>
- LIMA, J. S, PINTO, V. B, & FARIAS, M. G. (2020). O bibliotecário na gestão de dados de pesquisa: uma revisão sistemática. *Em Questão*, 26(3), 43-69. <<https://www.redalyc.org/articulo.oa?id=465664724003>>
- MONTEIRO, E. C. S. A., SANT'ANA, R. C. G. (2018). Plano de gerenciamento de dados em repositórios de dados de universidades. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, 23(53), 160-173. DOI: <<https://doi.org/10.5007/1518-2924.2018v23n53p160>>
- OCDE. OECD Principles and Guidelines for Access to Research Data from Public Funding. OECD, 2007. <<https://www.oecd.org/sti/inno/38500813.pdf>>

- PRADO, J. C., & MARZAL, M. A. (2013). Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents. *Libri*, 63(2), 123-131. <https://core.ac.uk/download/pdf/288499712.pdf>
- REIS, M. J. (2019). Ciência da Informação e Ciência de Dados: guia para alfabetização de dados para bibliotecários. [Dissertação de Mestrado Profissional, Universidade Federal de Sergipe]. Repositório da Universidade de Sergipe. [https://ri.ufs.br/bitstream/riufs/12667/2/MAKSON\\_%20JESUS\\_REIS.pdf](https://ri.ufs.br/bitstream/riufs/12667/2/MAKSON_%20JESUS_REIS.pdf)
- RICE, R., & SOUTHALL, J. (2016) The data librarian's handbook. In C. Ridsdale et al.(orgs.). *Strategies and Best Practices for Data Literacy Education Knowledge Synthesis Report*. Halifax, Canadá.
- SALES, L. F., & SAYÃO, L. F. (2019). Uma proposta de taxonomia para dados de pesquisa. *Conhecimento em Ação*, 4(1), 31-48. <https://revistas.ufrj.br/index.php/rca/article/view/26337>
- SANTA'ANNA, J., DIAS, C. C., & MACULAN, B. C. (2019). A gestão dos dados de pesquisa nas universidades e o papel dos serviços informacionais oferecidos nas bibliotecas: uma revisão narrativa. *Múltiplos Olhares em Ciência da Informação*, 9(2), 1-16. <https://periodicos.ufmg.br/index.php/moci/article/view/19169/16262>
- SAYÃO, L., & SALES, L. (2020). Afinal, o que é dado de pesquisa? *Biblos*, 34(2), 32-51. <https://periodicos.furg.br/biblos/article/view/11875>
- SEMELER, A. R., & PINTO, A. L. (2019). Os diferentes conceitos de dados de pesquisa na abordagem da biblioteconomia de dados. *Ciência da Informação*, 48(1). <http://hdl.handle.net/20.500.11959/brapci/111525>
- SHORISH, Y. (2015). Data Information Literacy and Undergraduates: a critical Competency. *College and Undergraduate Libraries*, 22(1), 97-106. DOI: <https://doi.org/10.1080/10691316.2015.1001246>
- SHINTAKU, M., APPEL, A. L., & OLIVEIRA, A. F. (2021). Tecnologias para gestão de dados de pesquisa segundo preceitos FAIR. In: L. F. Sales, V. S. Veiga, V. S. Oliveira, P. Henning, & L. F. Sayão. (orgs.). *Princípios FAIR aplicados à gestão de dados de pesquisa*. Rio de Janeiro: IBICT. <https://ridi.ibict.br/handle/123456789/1182>
- SILVEIRA, J. I. (2021). Acesso aberto a dados de pesquisa em repositórios universitários internacionais: um estudo sobre políticas de depósito, acesso e uso. (Dissertação de Mestrado, Universidade Federal do Rio Grande do Sul, Porto Alegre). <https://lume.ufrgs.br/handle/10183/222095>
- STEINEROVÁ, J., & ONDRIŠOVÁ, M. (2019). Research Data Literacy Perception and Practices in the Information Environment. In: S. Kurbanoglu et al. (orgs.). *Information Literacy in Everyday Life* (pp. 545-555) Finlândia:

Springer. <<https://link.springer.com/content/pdf/10.1007/978-3-030-13472-3.pdf>>

VITORINO, E. V. (2008). A formação contínua do profissional da informação: princípios epistemológicos à competência informacional. In: Anais IX Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação. (p. 15). São Paulo: USP. <<https://brapci.inf.br/index.php/res/download/181654>>

WHITMIRE, A. L., BOOCK, M., & SUTTON, S. C. (2015). Variability in academic research data management practices: Implications for data services development from a faculty survey. Program: *Electronic Library and Information Systems*, 49(4), 382-407. <<https://www.emerald.com/insight/content/doi/10.1108/PROG-02-2015-0017/full/html>>



# Criação de um chatbot para o ARCA - Repositório Institucional da Fiocruz

## **VALÉRIA MACHADO DA COSTA**

FUNDAÇÃO OSWALDO CRUZ

INSTITUTO DE COMUNICAÇÃO E INFORMAÇÃO CIENTÍFICA E TECNOLÓGICA EM SAÚDE

[valeria.machado@iciict.fiocruz.br](mailto:valeria.machado@iciict.fiocruz.br)

## **CLAUDETE FERNANDES DE QUEIROZ**

FUNDAÇÃO OSWALDO CRUZ

INSTITUTO DE COMUNICAÇÃO E INFORMAÇÃO CIENTÍFICA E TECNOLÓGICA EM SAÚDE

[claudete.queiroz@iciict.fiocruz.br](mailto:claudete.queiroz@iciict.fiocruz.br)

## **JOÃO PEDRO BORGES DA ROCHA GUIMARÃES**

FUNDAÇÃO OSWALDO CRUZ

INSTITUTO DE COMUNICAÇÃO E INFORMAÇÃO CIENTÍFICA E TECNOLÓGICA EM SAÚDE

[joao.borges@iciict.fiocruz.br](mailto:joao.borges@iciict.fiocruz.br)

## **CLARA GOMBERG FAULHABER DO VALE**

FUNDAÇÃO OSWALDO CRUZ

INSTITUTO DE COMUNICAÇÃO E INFORMAÇÃO CIENTÍFICA E TECNOLÓGICA EM SAÚDE

[clara.vale@iciict.fiocruz.br](mailto:clara.vale@iciict.fiocruz.br)

## **PALOMA C. MAMEDE**

FUNDAÇÃO OSWALDO CRUZ

INSTITUTO DE COMUNICAÇÃO E INFORMAÇÃO CIENTÍFICA E TECNOLÓGICA EM SAÚDE

[palomacmamede@gmail.com](mailto:palomacmamede@gmail.com)

## **ALINE DA SILVA ALVES**

FUNDAÇÃO OSWALDO CRUZ

INSTITUTO DE COMUNICAÇÃO E INFORMAÇÃO CIENTÍFICA E TECNOLÓGICA EM SAÚDE

[aline.alves@iciict.fiocruz.br](mailto:aline.alves@iciict.fiocruz.br)

## **CATARINA BARRETO MALHEIRO PEREIRA**

FUNDAÇÃO OSWALDO CRUZ

INSTITUTO DE COMUNICAÇÃO E INFORMAÇÃO CIENTÍFICA E TECNOLÓGICA EM SAÚDE

[catarina.barreto@icict.fiocruz.br](mailto:catarina.barreto@icict.fiocruz.br)

## **TATIANE MILITÃO**

UNIVERSIDADE FEDERAL FLUMINENSE

[tatimili2@yahoo.com.br](mailto:tatimili2@yahoo.com.br)

### **EIXO TEMÁTICO**

Comunicación académica, científica y cultural en abierto

### **RESUMO**

A Gestão do Relacionamento com o Cidadão tem como pressuposto atender o cidadão com rapidez, precisão e qualidade. Uma forma de se fazer isso é utilizando um assistente virtual, ou chatbot. O Arca é o repositório institucional da Fiocruz e, como tal, tem uma função estratégica tanto na instituição quanto fora, haja vista sua importância como o local de hospedagem e preservação de toda produção intelectual. A iniciativa começou com uma parceria entre o Centro de Tecnologia da Comunicação e da Informação (CTIC/Icict) e da equipe do Arca para melhorar a gestão com relação ao atendimento ao usuário, e também foi um primeiro passo para pensar em acessibilidade em um Repositório. Sendo assim, a criação de um chatbot para o ARCA, baseado em linguagem simples e acessibilidade, tem como objetivo aumentar a qualidade do atendimento ao cidadão, com um atendimento 24x7, obter estatísticas confiáveis, bem como otimizar o tempo de trabalho da equipe do repositório.

### **PALAVRAS CHAVE**

Chatbot; acessibilidade; Arca - Repositório institucional da Fiocruz.  
Chatbot; accessibility; Arca - Institutional Repository of Fiocruz.

## Introdução

O Citizen Relationship Management (CiRM), ou Gestão do Relacionamento com o Cidadão, é “uma filosofia gerencial que representa o esforço de todos os níveis de governo em atender com rapidez e precisão às necessidades, solicitações e indagações dos cidadãos, promovendo mudanças nas políticas, práticas e procedimentos do setor público” (Shaikh, Khan, 2014 apud Carvalho, Daniel et al.).

Aldo (2016) comenta que segundo a Política de Comunicação da Fiocruz que a Instituição

Precisa qualificar cada vez mais sua interação e comunicação direta com a população, por meio da valorização e integração dos serviços de relacionamento com o cidadão, além da interface com públicos específicos. Mais do que isso, o desafio não se resume a qualificar o atendimento, mas ampliar os canais de acesso e diálogo, e contribuir para a produção de conteúdos focada no cidadão, a partir da ausculta social, fortalecendo o caráter público, participativo e democrático da comunicação realizada pela Fundação (Pontes et al., 2016).

O Arca, Repositório Institucional da Fiocruz tem a missão de reunir, preservar e disponibilizar a produção intelectual da Instituição (Fundação Oswaldo Cruz, 2023). Para facilitar e atender as dúvidas dos seus usuários, foi implantada no ano de 2021 a ferramenta Fale Conosco, que demonstrou ser bastante eficaz para a equipe do Arca. Com o objetivo de otimizar ainda mais esse atendimento, foi proposta a criação de um chatbot.

O chatbot (do inglês chat - bate-papo e bot, de robot) é uma tecnologia que busca realizar tarefas dentro de uma determinada área do conhecimento, interagindo com as pessoas de modo semelhante a um humano (Calado, 2016). Seu funcionamento permite o esclarecimento de dúvidas do cidadão, 24 horas por dia, 7 dias por semana (24x7). Vale lembrar que a Lei 13.460, de 2017, afirma, em seu Art. 5º, inciso XIII, que “o usuário de serviço público tem direito à adequada prestação dos serviços, devendo os agentes públicos e prestadores de serviços públicos aplicarem soluções tecnológicas que visem a

simplificar processos e procedimentos de atendimento ao usuário e a propiciar melhores condições para o compartilhamento das informações” (Brasil, 2017).

O crescimento do uso de bots por diferentes empresas privadas favorece sua adoção pelo setor público (que já tem algumas iniciativas em andamento), uma vez que sua expansão aumenta as chances do cidadão já ter entrado em contato com este tipo de tecnologia, reduzindo seu estranhamento e sua curva de aprendizagem. Segundo a Pesquisa “Mapa do Ecosistema Brasileiro de Bots” (2020), em 2017 o Brasil tinha 8 mil bots criados. Em 2020 esse número saltou para 101 mil. Ainda de acordo com esta pesquisa, a finalidade mais comum dos bots criados em 2020 é a de atendimento (64%), isto é, a mesma área de atuação do chatbot do ARCA. Uma outra vantagem desta tecnologia é que ela funciona em diferentes plataformas (sites, redes sociais, whatsapp), o que permite diferentes abordagens de acordo com os públicos.

### **Justificativa**

Com o advento de novas tecnologias informáticas e a demanda de tornar o Repositório Arca mais ágil e acessível para os usuários, verificou-se a necessidade de implementar uma ferramenta que pudesse interagir com as pessoas. Nesse sentido, a escolha do chatbot foi para que o sistema conseguisse ter uma interface amigável, promovendo mais viabilidade e usabilidade em um ambiente disponibilizado na Web.

### **Metodología**

A metodologia do chatbot é separada em três etapas. A primeira consiste na construção do framework do bot. A segunda foi a elaboração do fluxo de conversação do chatbot e por último a implementação da acessibilidade.

Na construção do chatbot e pensando na sustentabilidade do chatbot foi utilizado o Rasa. É um framework de machine learning open source para

criação de chatbots (Dias, 2022). Por ser uma ferramenta open source, sua documentação está disponível na internet e sua comunidade oferece inúmeros exemplos e soluções que podem ser reaproveitados. Embora haja outros frameworks, como o Watson da IBM, que oferece uma interface mais amigável, quando falamos de serviço público, a questão do custo do produto e, conseqüentemente, sua sustentabilidade torna-se prioritária.

Definido o framework de construção do chatbot, iniciaram-se as discussões sobre a segunda parte da metodologia: o fluxo de conversação. Para a criação e revisão destes fluxos, optou-se pelo uso do MIRO, ferramentas online que é como um quadro branco infinito em que podemos acrescentar imagens, textos, setas, etc. Além de possuir uma versão gratuita que dá acesso a boa parte de suas funcionalidades, o MIRO foi escolhido por permitir: i) trabalho colaborativo; ii) uma curso de aprendizagem menor quando falamos de pessoas leigas atuando na construção do fluxo; iii) que as indicações de revisão sejam feitas na própria ferramenta, com comentários, prints de erros etc.; iv) poder ser utilizado para outras etapas do projeto, como planejamento e reuniões.

Para a construção do fluxo do diálogo, a primeira etapa foi ter acesso às dúvidas e respostas mais frequentes dos usuários em relação ao ARCA. A partir desta informação, montamos o fluxo do diálogo. A ideia foi criar um diálogo mais parecido possível com uma conversa humana, usando Natural Language Process (NLP).

### **Resultados preliminares**

Para este bot, optou-se por uma navegação guiada por menus, que significa que ele não permite que o usuário digite alguma pergunta. Todas as dúvidas foram organizadas nesse formato e o usuário vai clicando nos tópicos até encontrar a resposta. Ao final, perguntamos ao usuário se ele tem mais alguma dúvida. Caso responda NÃO, vai para nossa pesquisa de satisfação. Caso

resposta SIM, são apresentadas 3 opções: voltar ao menu do ARCA, conhecer os cursos da Fiocruz (já que este bot está integrado a um maior, que funciona no Portal Fiocruz), ou enviar uma mensagem para o Fale Conosco. Neste último caso, o cidadão sai do fluxo do bot e entra no fluxo de atendimento do Fale Conosco do ARCA. Esse último recurso é chamado de transbordo e também é oferecido quando o usuário avalia negativamente o chatbot, pois entendemos que desta forma é possível sanar a dúvida da pessoa pelo atendimento humano.

Com relação à pesquisa de satisfação, é importante salientar que, além do desenvolvimento da personagem Wal, um robzinho com a imagem da Fiocruz, foram criadas versões que representam emoções e buscam humanizar a interação entre o robô e o cidadão (Figura 1).



FIGURA 1. Versões do personagem Wal

Com o fluxo pronto, a próxima etapa foi a revisão do texto por uma especialista em UX Writer, que padronizou termos e deu o tom da conversação do chatbot. Também foi feita uma simplificação da linguagem, buscando simplificar, sempre que possível, termos mais complexos ou pouco conhecidos da população. Os textos também passaram por uma análise de legibilidade aplicando técnicas da linguagem simples e foram validados com a ajuda da ferramenta de legibilidade. O calculador de legibilidade fornece índices para

as seguintes métricas: Teste de facilidade de leitura de Flesch (Flesch reading ease), Índice Gulpease (Índice Gulpease), Nível de escolaridade de Flesch-Kincaid (Flesch-Kincaid grade level), Índice de nebulosidade de Gunning (Gunning fog index), Índice de leiturabilidade automatizado (Automated Readability Index - ARI) e Índice de Coleman-Liau (Coleman-Liau index) (SOUZA et al., 2021).

Em seguida, o fluxo foi validado com a equipe do Repositório Arca (Bibliotecários), para verificar se não havia nenhum erro. Importante destacar que o chatbot permite a inclusão de diferentes tipos de mídia.

No caso do Repositório, incluímos no fluxo do diálogo um vídeo tutorial. Com o fluxo aprovado, o chatbot foi implementado na área de homologação e testado pela equipe, visando encontrar algum erro de navegação, conteúdo ou mesmo de implementação online. Esta etapa durou aproximadamente 2 semanas, e após a correção dos erros detectados, o chatbot foi implementado na homepage principal ([www.arca.fiocruz.br](http://www.arca.fiocruz.br)), como mostra a Figura 2.

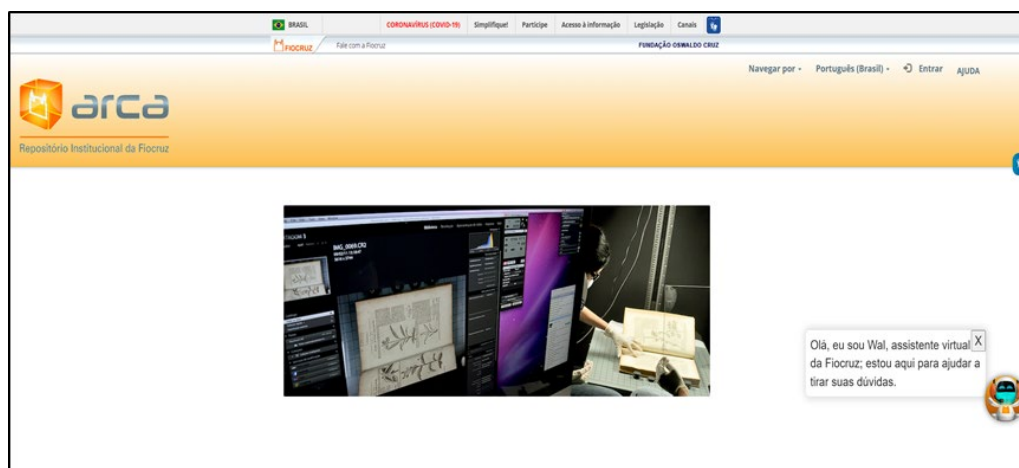


FIGURA 2. Chatbot na página principal do Arca

A implementação foi feita copiando um código direto no HTML da página, permitindo que o chatbot apareça sempre minimizado no canto inferior direito da tela. Para abri-lo, basta apenas clicar sobre ele que uma janela vai aparecer

com as questões sobre o Repositório de forma acessível a pessoas cegas que utilizam leitores de tela (Figura 3). Essa parte de acessibilidade inicia a terceira e última parte da metodologia de construção do bot.

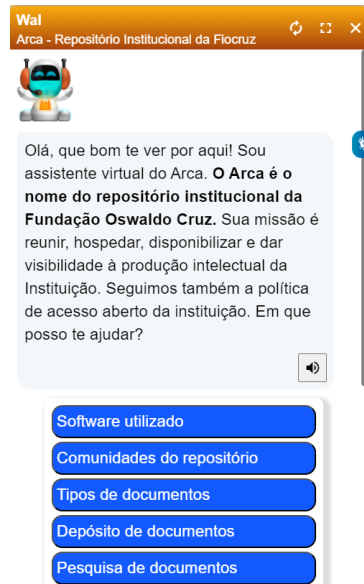


FIGURA 3. Chatbot na página principal do Arca Acessibilidade

Para ser acessível, o chatbot possui um designer de interfaces conversacional seguindo critérios de sucesso (A, AA e AAA) para alcançar as necessidades do máximo de usuários possível. Para alcançar as metas de um chatbot acessível, foram utilizadas as diretrizes propostas pelo Web Content Accessibility Guidelines (WCAG, 2022) e pelo Modelo de Acessibilidade em Governo Eletrônico (eMag) (BRASIL, 2022). As Páginas Web que não possuem um código de acordo com os padrões mencionados, apresentam comportamento imprevisível e, na maioria das vezes, impedem ou dificultam o acesso.

Em quatro meses de funcionamento, o chatbot do ARCA já totalizava 1.467 acessos únicos, com 74 transbordos, 78 avaliações positivas, 4 avaliações negativas e uma média semanal de 77 acessos únicos.



Em relação à pesquisa de satisfação, 97% das pessoas que avaliaram mostraram-se satisfeitas e 3% insatisfeitas. Importante destacar que após a pesquisa de satisfação é perguntado ao usuário se ele quer deixar algum comentário. Essa funcionalidade é muito importante pois nos permite entender melhor, em caso de avaliação negativa, qual foi o motivo da não satisfação do cidadão e, assim, melhorar o serviço seja na parte de conversação ou acessibilidade.

Embora o chatbot funcione de forma autônoma, é necessário que haja uma curadoria, de modo a identificar problemas e fazer as correções devidas. Para melhorar este serviço, foi criado um curso de curadoria, para a correção dos erros detectados na interação do bot com os usuários, seja erros de story (fluxo), de intent (assuntos) ou de conteúdo. O objetivo do curso é treinar as pessoas que ficaram responsáveis por supervisionar o funcionamento do chatbot.

Com relação à acessibilidade, simplificamos o fluxo conversacional do bot após estar no ar (texto explicativos, formato dos links e texto do menu), pois detectou-se, por meio de testes, que o leitor de tela não estava lendo corretamente as informações.

### **Considerações finais**

O número de chatbots no Brasil, em 2020, foi de 101 mil. Destes, somente 2% pertencem ao setor público. A criação do chatbot para o ARCA atendeu a algumas premissas do setor público: i) baixo custo, com uso preferencial de ferramentas gratuitas, o que permite a replicabilidade da solução para outras instituições públicas que possuam repositórios; ii) acessibilidade para pessoas com deficiência sensorial (cegas e surdas), garantindo o direito de acesso à informação; iii) atendimento com rapidez e precisão às necessidades, solicitações e indagações dos cidadãos; iv) melhoria na qualidade do atendimento ao cidadão (reforçada pela pesquisa de satisfação); e iv)

otimização do trabalho da equipe do repositório, que teve parte de sua carga horária liberada para dedicação em outras frentes.

## Bibliografía

- BRASIL. Lei nº 13.460. (2017). Dispõe sobre participação, proteção e defesa dos direitos do usuário dos serviços públicos da administração pública. Brasília, DF. <[http://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2017/lei/l13460.htm](http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2017/lei/l13460.htm)>
- BRASIL. Departamento de Governo Eletrônico. (2022). eMAG - Modelo de Acessibilidade em Governo Eletrônico. Brasília, DF. <<http://emag.governoeletronico.gov.br/>>
- CALADO, CAIO. (2016). Bots Brasil. O que é um chatbot? S. I. <<https://goo.gl/91wye4>>
- CARVALHO, DANIEL ET AL. (2021). Citizen Relationship Management (CiRM): the past, present, and future of an emerging concept. *Cadernos EBAPE.BR*, 19(1), 32-44. DOI: <<https://doi.org/10.1590/1679-395120200057x>>
- DIAS, TIAGO. (2022). Como Criar um Chatbot com Rasa Open Source. S. I.: Dados ao Cubo. <<https://dadosaocubo.com/como-criar-um-chatbot-com-rasa-open-source/>>
- FUNDAÇÃO OSWALDO CRUZ. (2023). Sobre o Arca. Rio de Janeiro. <<https://www.arca.fiocruz.br/terms/sobre.jsp>>
- MOBILE TIME. (2020). Mapa do ecossistema brasileiro de bots. S. I. <<https://www.mobiletime.com.br/pesquisas/mapa-do-ecossistema-brasileiro-de-bots-2020/>>
- PONTES, A. ET AL. (2016). Política de Comunicação da Fiocruz. Rio de Janeiro: Fiocruz/Presidência. 40 p. <<https://www.arca.fiocruz.br/handle/icict/41668>>
- SOUZA, M. P. M. DE ET AL. (2021). ALT: Análise de Legibilidade Textual. S. I. <<https://legibilidade.com/>>
- WCAG. W3C. (2022). Web Accessibility Initiative (WAI). S. I. <<https://www.w3.org/WAI/standards-guidelines/wcag/>>

# A aplicação da metodologia das boas práticas na submissão e autoarquivamento de recursos educacionais abertos em repositórios

**EVA PRISCILA VIEIRA DANN**

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

**CATERINA MARTA GROPOSO PAVÃO**

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

## **EIXO TEMÁTICO**

Comunicación académica, científica y cultural en abierto

## **RESUMO**

O presente artigo discute a respeito da adoção de boas práticas para a submissão e o autoarquivamento de recursos educacionais abertos em repositórios, de modo a incentivar a criação de políticas para o acervo desses objetos. Recursos educacionais abertos são materiais voltados ao ensino, aprendizagem e pesquisa, produzidos em diferentes suportes e formatos, depositados em repositórios digitais através do autoarquivamento, feito pelo próprio autor ou pessoa autorizada. Boas práticas são o conjunto de técnicas ou procedimentos identificados como os mais adequados para a realização de determinada tarefa. Com relação à metodologia empregada, trata-se de um estudo qualitativo, de natureza aplicada e com análise exploratório-descritiva. Inicialmente, realizou-se seleção de repositórios que pudessem ser avaliados quanto à adoção de boas práticas na avaliação dos objetos em tela. Após, procedeu-se à pesquisa documental nos repositórios selecionados e à análise dos dados obtidos. Com isso, foi possível compor um conjunto de boas práticas com aplicação possível para a submissão e autoarquivamento de recursos educacionais abertos, em diferentes contextos institucionais. Conclui-se que a elaboração de normativas para armazenar esses objetos é fundamental, pois proporciona melhoria dos fluxos e procedimentos e promove assim o compartilhamento, uso e reúso dos recursos educacionais disponibilizados.

## **PALAVRAS CHAVE**

Recurso educacional aberto; boas práticas; autoarquivamento. Open educational resource; best practices; self-archiving.

## **Introdução**

A Iniciativa dos Arquivos Abertos (Open Archives Initiative – OAI) e o Movimento do Acesso Aberto (Open Access Movement – OAM) são ações surgidas nos anos 1990 que visam armazenar, divulgar e compartilhar pesquisas científicas, de modo a trazer novas perspectivas aos dados, aos documentos e à informação, por meio das ferramentas livres de autoarquivamento (Galvino et al., 2020, p. 35). O acesso aberto (AA) propicia a criação de repositórios digitais, instrumentos concebidos com a intenção de facilitar o acesso, organizar, reunir e disponibilizar a produção científica dos pesquisadores, podendo dividir-se em repositórios temáticos ou institucionais (Leite et al., 2012, p. 7).

Assim, os repositórios digitais são bases de dados de fonte, voltadas à divulgação científica (Rocha; Dorfman, 2015), sendo através deles que se disponibilizam os recursos informacionais em formato digital. Nesse sentido, podemos considerar os repositórios institucionais (RIs) como elementos essenciais ao AA, pois eles são elementos de inovação do sistema de comunicação científica e dos meios de gerenciamento da informação (Leite et al., 2012, p. 7). Os RIs são grandes bibliotecas digitais, onde o conhecimento de uma coletividade é armazenado, aspirando, como já dito, o acesso e a preservação de longo prazo. Por isso, podemos dizer que os RIs possuem uma função também política, já que eles são capazes de integrar a rede informacional de um país, constituindo-se em atores políticos do movimento internacional de apoio ao livre acesso à informação científica (Marcondes; Sayão, 2009, p. 9).

Atualmente, em virtude das mudanças ocorridas nos modos de ensinar e aprender que permitiu a inserção das Tecnologias da Informação e Comunicação (TICs) na Educação, os RIs armazenam, entre uma gama de itens, os Recursos Educacionais Abertos (REAs). Em virtude de suas características e finalidades, os RIs são instrumentos que favorecem o uso e reúso dos REAs, no âmbito das universidades e institutos de pesquisa, uma vez que possibilitam armazenar, preservar e disseminar os materiais neles depositados. Por objetos digitais com características e finalidades próprias, cabe a essas instituições desenvolver políticas e promover o desenvolvimento de boas práticas para a submissão e autoarquivamento de REAs em seus RIs, no sentido de fomentar não só a criação, mas também a publicação de REAs entre seus pesquisadores.

Este trabalho tem o objetivo de refletir sobre a aplicação de boas práticas aos procedimentos de submissão e autoarquivamento de REAs em repositórios. Para tanto, está dividido da seguinte maneira: na seção 2, delimitamos as definições e características dos REAs e da educação aberta (EA). Na seção 3, descrevemos o conceito de boas práticas e abordamos as políticas de coleção que, de modo geral, sustentam os procedimentos de submissão e autoarquivamento de itens em repositórios. A seção 4 expõe a metodologia que elaboramos para, a partir da seleção e pesquisa documental em RIs brasileiros e americanos, identificar e elencar um conjunto de boas práticas para a submissão e autoarquivamento de REAs. A seção 5 apresenta os resultados preliminares encontrados. Na última seção, encontram-se as nossas considerações finais.

### **A educação aberta e os REAs**

A evolução dos artefatos comunicacionais e o aparecimento de novas tecnologias promoveram uma série de transformações com impactos marcantes sobre as relações sociais, as quais mudaram a sociedade e as

formas de aquisição do conhecimento. Tais transformações, especialmente as ocorridas ao longo do século XX, foram fundamentais para a formação da sociedade em rede, produto da sociedade da informação.

Com relação à Educação, Cardoso e Pinto (2021, p. 78-79) afirmam que a realidade presente na sociedade em rede vem alterando processos educativos e contribuindo com a recuperação da antiga dimensão comunitária de aprendizagem, o que permite a ascensão de novos meios de autoformação. Com isso, temos um deslocamento no modo como pensamos a função social da educação. Essa conjuntura cria as condições necessárias para o advento da EA. Compreendida como um conceito mais amplo que o da Educação a Distância (EaD), a EA surgiu nos anos 1970 como um dos movimentos educacionais mais relevantes do século XXI.

Tendo por base, entre outros, a convergência e evolução dos recursos educativos abertos, do software livre, do livre acesso, dos MOOCs, da ciência aberta e de um conjunto de mudanças sociais e econômicas, a matriz deste movimento ultrapassa o mero acesso a conteúdos e recursos e associa-se a uma nova filosofia educativa, a novos valores baseados na abertura, na ética da participação e na colaboração (Aires, 2016, p. 258).

A EA consiste em um modelo educacional que utiliza as TICs nos processos de ensino-aprendizagem, a fim de tornar o conhecimento acessível e gratuito a todos, flexibilizar processos avaliativos, entre outros aspectos, podendo ser aplicada a diferentes modalidades e níveis de ensino. Ela configura uma série de operações, as quais objetivam democratizar o ensino e tornar o conhecimento disponível a todo e qualquer cidadão.

Com propósitos semelhantes, os REAs são dispositivos que impulsionam mudanças nos processos de ensino-aprendizagem. Amiel (2012, p. 24) nos diz que o acesso a esses materiais é essencial para a criação de arranjos de ensino e aprendizagem mais versáteis. O autor compreende que os REAs colaboram na formulação de novas configurações de aprendizagem e ensino, sendo caracterizados como todo e qualquer material produzido com fins

pedagógicos, desde que sejam de domínio público e estejam disponibilizados com uma licença aberta. Além disso, os REAs, da mesma forma que os movimentos do AA e da EA, pertencem ao ecossistema da CA e foram fortemente influenciados pelo desenvolvimento das TICs.

De acordo com Dutra e Tarouco (2007, p. 2), REAs são “[...] materiais educacionais digitais disponibilizados de forma livre e aberta para a comunidade acadêmica em geral, que os utilizam para o ensino, aprendizagem e pesquisa.” A primeira definição mais conhecida e amplamente difundida foi elaborada pela UNESCO em 2002, onde considera que os REAs são “[...] materiais de ensino, aprendizado e pesquisa em qualquer meio disponível no domínio público, que foram disponibilizados com licenças abertas, permitindo acesso, uso, redesignação, reutilização e redistribuição por terceiros, com poucas ou sem nenhuma restrição [...]” (United..., 2011). Assim, um REA pode abranger diferentes tipos, suportes e conteúdos de aprendizagem, como cursos completos; módulos de cursos; livros; materiais de disciplinas; artigos; vídeos; áudios; podcasts; animações; simulações; sites; jogos; mapas; imagens; infográficos; entre outros objetos de aprendizagem.

A fim de facilitar seu uso e reúso, um REA necessita distinguir-se por aspectos técnicos e legais. Dessa forma, eles estão baseados nos princípios da abertura técnica, garantida pela interoperabilidade, e das licenças de uso, que asseguram aos autores o crédito da obra, permitindo, simultaneamente, que os usuários do REA possam copiá-lo, compartilhá-lo e distribuí-lo (Educação..., 2013). Uma das licenças abertas amplamente aceitas e utilizadas são as Creative Commons (CC), as quais permitem socializar o conhecimento produzido, sem tirar o mérito do autor (Pavão et al., 2020, p. 236).

Para Mazzardo et al. (2020), são as permissões dos REAs as responsáveis por informar o que é possível fazer com o recurso. Os autores citam que, através das liberdades do software livre, Willey definiu os chamados 5R para abertura dos REAs, capazes de ampliar as possibilidades de uso pedagógico de tais materiais. A viabilização dos 5R de abertura dos REAs ocorre justamente

em virtude das licenças abertas. Além disso, consistem em uma maneira de aumentar a produção e o compartilhamento de REAs e, conseqüentemente, o acesso ao conhecimento. A Figura 1 apresenta os 5R de abertura dos REAs.



**Fonte:** (FENGCHUN et al., 2019; FURNIEL; MENDONÇA; SILVA, 2020, adaptado).

FIGURA 1. 5R de abertura dos REAs

Fonte: Gusmão et al. (2020, p. 8).

As noções estabelecidas aos REAs admitem certa flexibilidade, já que instituições e órgãos de pesquisa produtores desses objetos necessitam conceituá-los, de tal modo que o conceito estabelecido faça sentido aos propósitos e à realidade da instituição. Logo, cabe compreender que os REAs precisam possuir características comuns a toda a comunidade científica, enquanto os conceitos mais específicos são delimitados pelas próprias instituições. De acordo com a UNESCO:

Independentemente de sua definição conceitual, os REA abrangem uma ampla gama de recursos de aprendizagem. As características gerais dos REA são as seguintes:



- Podem designar qualquer tipo de recurso de aprendizagem.
- Frequentemente, mas não exclusivamente, são apresentados em formato digital.
- formato facilita a reutilização, o intercâmbio e a adaptação do recurso a um ambiente educacional diferente do original, por isso são muitas vezes digitais (United..., 2020, p. 12, tradução nossa).

Portanto, elaborar definições claras aos REAs, alinhadas ao planejamento estratégico da instituição é, sem dúvida, um dos motores de fomento ao autoarquivamento de REAs em repositórios. Uma comunidade acadêmico-científica que sabe efetivamente o que são recursos educacionais, que dispõe de normativas e orientações sobre quais são as características e os requisitos pedagógicos, técnicos e legais envolvidos, certamente terá maior engajamento na produção, uso e reúso desses objetos.

### **As boas práticas e as políticas de submissão e autoarquivamento em repositórios**

A metodologia conhecida por boas práticas é uma abordagem de aplicação em diversas áreas do conhecimento. Com origem na expressão inglesa *best practices* ou *good practices*, as boas práticas são um conjunto de técnicas ou procedimentos identificados como os mais adequados para a realização de determinada tarefa. Elas possuem o intento de arrolar requisitos ou padrões de gestão, proporcionando melhores resultados. Bretschneider et al. (2005), assim as definem:

O termo “boas práticas” implica aquilo que é o melhor, quando comparado a qualquer alternativa para uma ação em curso e aquilo que é uma prática planejada, para alcançar algum objetivo. Por isso, existem três características importantes associadas a uma “boa prática”:

- 1) um processo comparativo,
- 2) uma ação, e

- 3) uma ligação entre a ação e algum resultado ou meta. (Bretschneider et al., 2005, p. 309, tradução nossa).

Por estabelecer requisitos mínimos e protocolos para a realização de tarefas, as boas práticas, aplicadas aos repositórios tanto na fase de construção quanto na de manutenção, asseguram sua qualidade, oportunizando maior visibilidade (Leite et al., 2012). Entendemos que, dentre as boas práticas aplicáveis aos repositórios, estão as que se referem à implementação de políticas de autoarquivamento de itens.

O OAM é um dos grandes impulsionadores das políticas de autoarquivamento, sendo esta, conforme Pinheiro e Ferrez (2014, p. 36) e Freitas et al. (2021, p. 170), a ferramenta resultante dessas iniciativas. O autoarquivamento consiste no depósito de um documento pelo próprio autor em um repositório digital. Para tanto, são utilizados conjuntos de metadados definidos pela Open Archives Initiative, organização norte-americana que desenvolve e promove padrões de interoperabilidade, o que favorece o intercâmbio do conteúdo e de informações em formatos digitais.

A submissão é o procedimento relacionado à etapa de entrega da versão final de determinado item, para depósito em um repositório. Ela envolve a transferência de custódia do material do produtor ao consumidor. De acordo com Flores et al. (2017), o modelo de referência Open Archival Information System (OAIS) engloba os conceitos mais relevantes para o arquivamento de objetos digitais, definindo “[...] um ambiente de preservação digital, por meio de um conjunto mínimo de responsabilidades [...]” (Flores et al., 2017, p. 74). O modelo traz, ainda, a definição de três papéis para as entidades de seu ambiente, quais sejam, produtor, administrador e consumidor.

O produtor é o papel desempenhado por pessoas ou sistemas externos que fornecem a informação a ser preservada. Nesse caso, os produtores podem ser tantos os sistemas de produção de documentos digitais quanto os indivíduos que estão diretamente ligados a sua produção. O Administrador é o papel

desempenhado por aqueles que definem a política em um nível mais amplo, como parte de uma organização. Como, por exemplo, a política de gestão e de preservação digital, que deve ser desenvolvida pelo arquivo com o apoio da alta administração a fim de ser a principal fonte de financiamento de um OAIS. No entanto, não está envolvida diretamente nas operações de arquivo do dia-a-dia. O consumidor é o papel desempenhado por pessoas que procuram encontrar informações de seu interesse, através de pedido de informações ou por plataformas que viabilizam o acesso aos documentos (Flores et al., 2017, p. 74).

Com relação à submissão, o OAIS indica que ela ocorre por meio da entidade funcional denominada ingestão, onde o pacote de submissão é avaliado e conferido, podendo existir para tanto um fluxo de submissão. Tal entidade é composta pelos “[...] serviços necessários para preparar os conteúdos de armazenamento e gerenciamento de dados dentro do Arquivo OAIS, através de Pacotes de Informação para Submissão (SIP).” (Flores et al., 2017, p. 74). Em vista dos elementos apontados, acreditamos que o incentivo à implantação de políticas de submissão e autoarquivamento deve ser continuamente estimulado nas universidades e institutos de pesquisa, com vistas a conduzir os autores e pesquisadores ao depósito dos objetos digitais que são foco deste trabalho.

### **Procedimentos metodológicos**

Este estudo consiste em uma pesquisa qualitativa de natureza aplicada, com análise exploratório-descritiva, que se encontra em fase de ampliação de suas investigações. No projeto-piloto, realizado no período compreendido entre novembro/2022 a março/2023, procedemos inicialmente a uma seleção de RIs, com o intuito de localizar aqueles que pudessem ser analisados, quanto à adoção de boas práticas na elaboração e avaliação de REAs. Para tanto, realizamos buscas por repositórios com boa visibilidade, através do Ranking

Web of Repositories, o Webometrics. Os registros de ranqueamento utilizados foram os da 14ª edição do Transparent Ranking: Institutional Repositories by Google Scholar, datada de junho de 2022. A fim de compor uma amostra, optamos por escolher os três primeiros repositórios brasileiros e os três primeiros americanos que contivessem os parâmetros desejados, isto é, uso do software Dspace e ser de um repositório do tipo institucional.

A partir disso, buscamos o repositório no Webometrics, conforme o ranqueamento e, em seguida, por meio do Directory of Open Access Repositories (OpenDOAR), conferimos se os parâmetros estavam presentes. Após, procedemos ao acesso à URL do repositório, com o intuito de localizar explicitamente os REAs em cada um deles, procurando pelo tipo de documento/conteúdo ou comunidade/coleção. Decidimos, ainda, selecionar somente aqueles repositórios que destinam um espaço específico aos REAs, podendo ser tal espaço uma comunidade, subcomunidade ou coleção. Os repositórios brasileiros encontrados sob esta perspectiva foram o Repositório Digital Institucional da Universidade Federal do Paraná – Acervo Digital; o Repositório Institucional Fundação Oswaldo Cruz – ARCA; e o Repositório da Universidade Federal de Pernambuco – Attena. Com relação aos repositórios americanos, foram selecionados o MIT Institutional Repository, o University of Michigan Deep Blue e o OAK Trust Texas A&M University Repository. Cabe ressaltar que o Repositório Digital da UFRGS, o Lume, que ocupava o segundo lugar entre os RIs mais bem colocados na 14ª edição do Transparent Ranking foi excluído da seleção, já que tínhamos o intuito de observar outras realidades, distintas do nosso contexto institucional.

A etapa seguinte configurou-se na realização de pesquisa documental nos sítios de cada repositório, de modo a localizar instrumentos, manuais, tutoriais, treinamentos ou outros materiais, que demonstrassem orientações de caráter institucional dos repositórios aos autores/depositantes, com vistas à divulgação de boas práticas entre suas comunidades acadêmicas, com enfoque na submissão e no autoarquivamento de REAs. Por fim, as

informações obtidas foram analisadas e, a partir delas, foi possível compor um conjunto de boas práticas adotadas pelos RIs examinados. Na próxima seção, listamos as boas práticas que acreditamos ser de aplicação possível ao contexto dos RIs em diferentes cenários e instituições.

### **Resultados preliminares**

Os repositórios analisados apresentam duas importantes particularidades. Com relação à primeira, podemos dizer que enquanto os RIs brasileiros dispõem de normativas ou orientações específicas para o depósito de REAs, os RIs americanos não as disponibilizam de modo específico, mas contemplando todas as suas comunidades, subcomunidades e/ou coleções.

No que concerne à segunda diferenciação, embora os RIs americanos não contenham normativas exclusivas aos REAs, estas, em algum momento, versam sobre a existência da figura de um administrador ou grupo de administradores para suas comunidades, subcomunidades e coleções. Além de nos fazer convergir para a ideia de que a avaliação por pares em RIs está melhor consolidada nos Estados Unidos, quando comparada ao Brasil, o fato vai de encontro ao que compreendemos ser a tarefa de um comitê consultivo para REAs.

Em suma, a partir da pesquisa documental realizada e percebendo que RIs com normas e políticas adequadas possuem comunidades, estruturas e arranjos mais consolidados nas instituições de ensino e pesquisa a que estão ligados, conseguimos listar um conjunto de boas práticas, que julgamos relevantes de serem observadas pelas equipes de trabalho durante os procedimentos de submissão, depósito e armazenamento de REAs:

- 1)** definir/conceituar REA em nível institucional, pois é fundamental a presença de definições claras, que informem o que é e o que não é REA para a instituição;

- 2) prever uma política para os acervos dos REAs, que oriente e dê ciência acerca dos procedimentos que a comunidade acadêmica precisa seguir na inclusão de conteúdos.
- 3) quando os procedimentos forem realizados por mais de uma equipe responsável, estas devem elaborar checklists e manuais de procedimentos, preferencialmente em conjunto ou partilhar informações entre si;
- 4) divulgar para sua comunidade e público em geral sobre os benefícios da produção, publicação e disseminação de REAs, fomentando o uso e reuso dos objetos;
- 5) determinar local(is) específico(s) de seu repositório para o armazenamento dos REAs, indicando se estes objetos serão subdivididos de alguma maneira (por exemplo, por área do conhecimento) ou não;
- 6) orientar seus autores quanto às características técnicas e pedagógicas dos REAs. Nesse sentido, é importante a construção de tutoriais para autores e, quando o depósito é mediado, também para depositantes e revisores/validadores;
- 7) indicar as tipologias possíveis ao autoarquivamento de REAs no RI, os formatos de arquivo admitidos para cada tipologia, as orientações gerais para o upload de arquivos, os critérios e formas de embargo de conteúdos, entre outros;
- 8) indicar sob que licenças os REAs são disponibilizados;
- 9) orientar rigorosamente quais metadados e como os metadados de um RE precisam ser descritos, por que são descritos de determinada forma e qual a importância da descrição correta;
- 10) ter um administrador (pessoa ou grupo) para cada comunidade, subcomunidade ou coleção do RI que contenha REAs, para que melhor gestão técnica do acervo, e um ou mais administradores, com funções exclusivas de curadoria dos conteúdos, garantindo assim o cumprimento de requisitos técnicos e pedagógicos.

## Considerações finais

As políticas existentes em repositórios devem favorecer a efetivação, manutenção adequada e revisão de métodos, de modo a promover o engajamento da comunidade acadêmica a qual pertence. De acordo com Melero (2007), o sucesso do autoarquivamento em repositórios institucionais deve-se ao fato de que estes são realmente a imagem da produção científica da instituição. Para o autor, as políticas sobre exigência ou recomendação exercem grande influência na efetividade do autoarquivamento dos objetos.

Neste trabalho, intentamos elencar, a partir da seleção e análise de RIs brasileiros e americanos, um conjunto de boas práticas aplicáveis na elaboração, submissão e autoarquivamento de REAs em repositórios. Entendemos que a pesquisa aqui apresentada pode qualificar os fluxos e procedimentos adotados na submissão e autoarquivamento de REAs. No âmbito das universidades e órgãos de pesquisa, a adoção de boas práticas pode promover debates institucionais sobre os REAs, itens depositados nos acervos com características e finalidades peculiares. Como sequência aos estudos, pretendemos ampliar a pesquisa para mais instituições, além de outros países, o que provavelmente aprimorará os resultados.

## Bibliografía

- ACERVO DIGITAL DA UFPR. (2022). [Portal eletrônico da UFPR].  
<<https://acervodigital.ufpr.br/>>
- AIRES, L. (2016). E-Learning, educação online e educação aberta: contributos para uma reflexão teórica. *RIED*, 19(1): 253-269.  
<<https://repositorioaberto.uab.pt/bitstream/10400.2/5034/1/14356-27074-1-PB.pdf>>
- AMIEL, T. (2012). Educação aberta: configurando ambientes, práticas e recursos educacionais. In: Santana, Bianca; Rossini, Carolina; Pretto, Nelson De Lucca (Org.). *Recursos Educacionais Abertos: práticas colaborativas e políticas públicas* (pp. 17-33).  
<<https://www.aberta.org.br/livrorea/livro/livroREA-1edicao-mai2012.pdf>>

- ARCA, REPOSITÓRIO INSTITUCIONAL DA FIOCRUZ. (2022). [Portal eletrônico da Fiocruz]. <<https://www.arca.fiocruz.br/>>
- ATTENA, REPOSITÓRIO DIGITAL DA UFPE. (2022). [Portal eletrônico da UFPE]. <<https://attena.ufpe.br/>>
- BRETSCHNEIDER, S.; MARC-AURELE JUNIOR, F. J.; WU, J. (2005). Best practices research: a methodological guide for the perplexed. *Journal of Public Administration Research and Theory*, 15(2), 307-323. DOI: <<https://doi.org/10.1093/jopart/mui017>>
- CARDOSO, T. M. L.; PINTO, J. P. (2021). Recursos educacionais abertos e educação aberta na sociedade em rede. *Páginas A&B, n. esp.*, 78-82. <<http://hdl.handle.net/20.500.11959/brapci/157330>>
- CYBERMETRICS LAB. (2022). Ranking Web of Repositories. Transparent ranking: institutional repositories by Google Scholar. 14th. ed. <<https://repositories.webometrics.info/en/institutional>>
- DEEP BLUE DOCUMENTS. (2022). [Portal eletrônico da University of Michigan Library]. <<https://deepblue.lib.umich.edu/documents>>
- DSPACE@MIT. (2022). [Portal eletrônico do Massachusetts Institute of Technology]. <<https://dspace.mit.edu/>>
- DUTRA, R. L. DE S.; TAROUÇO, L. M. R. (2007). Recursos educacionais abertos (Open Educational Resources). *Revista Novas Tecnologias na Educação, UFRGS*, 5(1). <<http://seer.ufrgs.br/index.php/renote/article/view/14171/8099>>
- EDUCAÇÃO ABERTA. (2013) Recursos Educacionais Abertos (REA): um caderno para professores. <<http://educacaoaberta.org/cadernorea>>
- FLORES, D.; PRADEBON, D. S.; CÉ, G. (2017). Análise do conhecimento teórico-metodológico da preservação digital sob a ótica da OAIS, SAAI, ISO 14721 e NBR 15472. *Brazilian Journal of Information Science: Research Trends*, 11(4), 72-80. <<https://revistas.marilia.unesp.br/index.php/bjis/article/view/7511/4790>>
- FREITAS, M. P. DE; DAL'EVEDOVE, P. R.; TARTAROTTI, R. C. D. (2021). Políticas de autoarquivamento em repositórios institucionais brasileiros: estudo analítico do metadado assunto. *Páginas A&B, n. esp.*, 169-175. <<https://brapci.inf.br/index.php/res/v/157369>>
- GALVINO, C. C. T.; ROSA, M. N. B.; OLIVEIRA, B. M. J. F. DE. (2020). O movimento de Acesso Aberto e a Ciência Aberta: uma proposta de repositório de dados e memória na universidade federal de alagoas. *Ciência da Informação em Revista*, 7(1), 34-45. <<https://brapci.inf.br/index.php/res/download/139970>>
- GUSMÃO, C. M. G. DE; MACHIAVELLI, J. L.; SILVA, A. C. V.; PONTES, S. V. N. S.; CORREIA, J. A. DA S.; BORBA, V. DA R. (2020). Manual para autodepósito de Recursos Educacionais Abertos no repositório attena da Universidade Federal de Pernambuco. <<https://repositorio.ufpe.br/handle/123456789/38464>>



- LEITE, F. C. L.; AMARO, B.; BATISTA, T.; COSTA, M. (2012). Boas práticas para a construção de repositórios institucionais da produção científica. Rio de Janeiro: IBICT. <<https://livroaberto.ibict.br/handle/1/703>>
- MARCONDES, C. H.; SAYÃO, L. F. Introdução: repositórios institucionais e livre acesso. In: Sayão, L. F.; Toutain, L. B.; Rosa, F. G.; Marcondes, C. H. (org.). (2009). Implantação e gestão de repositórios institucionais: políticas, memória, livre acesso e preservação (pp. 9-21). Salvador: EDUFBA. <<https://repositorio.ufba.br/handle/ufba/473>>
- MAZZARDO, M. D.; MALLMANN, E. M.; JACQUES, J. S.; SCHNEIDER, D. DA R.; SCHRAIBER, R. T.; LAUERMANN, R. A. C.; ALBERTI, T. F.; MORISSO, M. M.; REGINATTO, A. A. (2020). Fluência Tecnológico-Pedagógico (FTP) e recursos educacionais abertos (REA). Santa Maria: UFSM, GEPETER. <<https://gepeter.proj.ufsm.br/pressbook/livrorea/>>
- MELERO, R. (2007). Políticas sobre el libre acceso a la producción científica y la respuesta de los autores. In: IV Congreso de Comunicación Social de la Ciencia, Madrid, Espanha. <<https://digital.csic.es/bitstream/10261/1492/1/OA5rm.pdf>>
- OPENDOAR. (2022). <<https://v2.sherpa.ac.uk/opensoar/>>
- PAVÃO, C. G.; CARNEIRO, M. L. F.; FERREIRA, M. K. (2020). A implantação da comunidade de recursos educacionais no Lume/UFRGS. In: Mallmann, E. M.; Jacques, J. S.; Reginatto, A. A.; Alberti, T. F. (Orgs.). REA: teoria e prática (pp 229-246). Pimenta Cultural. <<https://lume.ufrgs.br/handle/10183/217355>>
- PINHEIRO, L. V. R.; FERREZ, H. D. (2014). Tesouro Brasileiro de Ciência da Informação. Rio de Janeiro: IBICT. <<http://sitehistorico.ibict.br/publicacoes-e-institucionais/tesouro-brasileiro-de-ciencia-da-informacao-1>>
- ROCHA, R. P. DA; DORFMAN, A. (2015). O Unbral Fronteiras em busca de boas práticas na divulgação e preservação da produção acadêmica. *Anuário Unbral das fronteiras brasileiras*, 1, 87-91. <<https://lume.ufrgs.br/handle/10183/150003>>
- THE OAKTRUST DIGITAL REPOSITORY. (2022). [Portal eletrônico da Texas A&M University Libraries]. <<http://oaktrust.library.tamu.edu/>>
- UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION. (2020). Directrices para la elaboración de políticas de recursos educativos abiertos. Paris: UNESCO. <<https://unesdoc.unesco.org/ark:/48223/pf0000373558>>
- UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION. (2011). Guidelines for open educational resources (OER) in higher education. France: UNESCO. <<https://unesdoc.unesco.org/ark:/48223/pf0000213605>>

# Autoavaliação de confiabilidade do Arca – Repositório Institucional da Fiocruz pela norma ISO-16363:2012: apresentação dos resultados consolidados

## **CATARINA BARRETO MALHEIRO PEREIRA**

FUNDAÇÃO OSWALDO CRUZ

INSTITUTO DE COMUNICAÇÃO E INFORMAÇÃO CIENTÍFICA E TECNOLÓGICA EM SAÚDE

[catarina.barreto@iciict.fiocruz.br](mailto:catarina.barreto@iciict.fiocruz.br)

## **CLAUDETE FERNANDES DE QUEIROZ**

FUNDAÇÃO OSWALDO CRUZ

INSTITUTO DE COMUNICAÇÃO E INFORMAÇÃO CIENTÍFICA E TECNOLÓGICA EM SAÚDE

[claudete.queiroz@iciict.fiocruz.br](mailto:claudete.queiroz@iciict.fiocruz.br)

## **LUCIANA DANIELLI DE ARAUJO**

FUNDAÇÃO OSWALDO CRUZ

INSTITUTO DE COMUNICAÇÃO E INFORMAÇÃO CIENTÍFICA E TECNOLÓGICA EM SAÚDE

[luciana.danielli@iciict.fiocruz.br](mailto:luciana.danielli@iciict.fiocruz.br)

## **RAPHAEL BELCHIOR RODRIGUES**

FUNDAÇÃO OSWALDO CRUZ

INSTITUTO DE COMUNICAÇÃO E INFORMAÇÃO CIENTÍFICA E TECNOLÓGICA EM SAÚDE

[raphael.rodrigues@iciict.fiocruz.br](mailto:raphael.rodrigues@iciict.fiocruz.br)

## **ALINE ALVES DA SILVA**

FUNDAÇÃO OSWALDO CRUZ

INSTITUTO DE COMUNICAÇÃO E INFORMAÇÃO CIENTÍFICA E TECNOLÓGICA EM SAÚDE

[aline.alves@iciict.fiocruz.br](mailto:aline.alves@iciict.fiocruz.br)

### **RITA DE CASSIA DA SILVA**

FUNDAÇÃO OSWALDO CRUZ

INSTITUTO DE COMUNICAÇÃO E INFORMAÇÃO CIENTÍFICA E TECNOLÓGICA EM SAÚDE

[rita.cassia@icict.fiocruz.br](mailto:rita.cassia@icict.fiocruz.br)

### **IGOR FALCE DIAS DE LIMA**

FUNDAÇÃO OSWALDO CRUZ

INSTITUTO DE COMUNICAÇÃO E INFORMAÇÃO CIENTÍFICA E TECNOLÓGICA EM SAÚDE

[igor.lima@icict.fiocruz.br](mailto:igor.lima@icict.fiocruz.br)

### **MARIA DA CONCEIÇÃO CALMON ARRUDA**

FUNDAÇÃO OSWALDO CRUZ

INSTITUTO FERNANDES FIGUEIRA DA MULHER, DA CRIANÇA E DO ADOLESCENTE

[maria.calmon@icict.fiocruz.br](mailto:maria.calmon@icict.fiocruz.br)

### **ALEX HOLANDA**

ARQUIVO NACIONAL

[alex.holanda@gmail.com](mailto:alex.holanda@gmail.com)

### **TIAGO MARTINS DA COSTA FERREIRA**

FUNDAÇÃO OSWALDO CRUZ

INSTITUTO DE COMUNICAÇÃO E INFORMAÇÃO CIENTÍFICA E TECNOLÓGICA EM SAÚDE

[tiago.costa@icict.fiocruz.br](mailto:tiago.costa@icict.fiocruz.br)

### **EIXO TEMÁTICO**

Infraestrutura tecnológica

### **RESUMO**

Apresenta o percurso de autoavaliação do Arca – Repositório Institucional da Fiocruz, iniciado em 2019, visando recomendar as ações necessárias para atender os critérios e requisitos descritos na norma ISO 16363:2012. Para dar continuidade ao trabalho, foi submetido um Projeto na Fiocruz para efetuar uma nova autoavaliação, visando apresentar os resultados

obtidos em 2023 e fazer uma análise com os que foram mensurados em 2019, para definir como se encontra a confiabilidade do Repositório, como também atualizar o Plano de Preservação Digital. A metodologia utilizada foi construída a partir das questões formuladas e respondidas conforme os requisitos recomendados na Norma, utilizando a escala do tipo Likert para medir o nível de adesão e atribuindo um grau (de 0 a 4). Como resultado, obtivemos um novo mapeamento dos dados relacionados aos critérios Infraestrutura Organizacional, Gestão dos Objetos Digitais, Gestão de Risco de Infraestrutura e Segurança, com a identificação dos pontos fortes e fracos e riscos diagnosticados. O trabalho serviu para implementar as correções dos requisitos detectados como incoerentes ou não realizados. O projeto contou com a equipe do Repositório Arca, do Arquivo Nacional, da Biblioteca de Manguinhos e da Biblioteca da Mulher, da Criança e do Adolescente, ambas da Fiocruz.

#### **PALAVRAS CHAVES**

Repositórios digitais confiáveis; certificação; confiabilidade; Arca – Repositório Institucional da Fiocruz; preservação digital. Trusted digital repositories; certification; reliability; Arca – Institutional Repository of Fiocruz; digital preservation.

#### **Introdução**

A **Fundação Oswaldo Cruz**, instituição pública vinculada ao Ministério da Saúde (MS) no Brasil, é referência mundial na produção tecnológica de serviços e insumos estratégicos para a promoção da saúde da população, a consolidação e o fortalecimento do Sistema Único de Saúde (SUS). Em 2020, a Fundação teve um papel fundamental e importante no enfrentamento da Emergência de Saúde Pública decorrente do Novo Coronavírus (Covid-19). Sua “cadeia de valores” envolve diferentes áreas e processos, conforme apresentado na Figura 1.

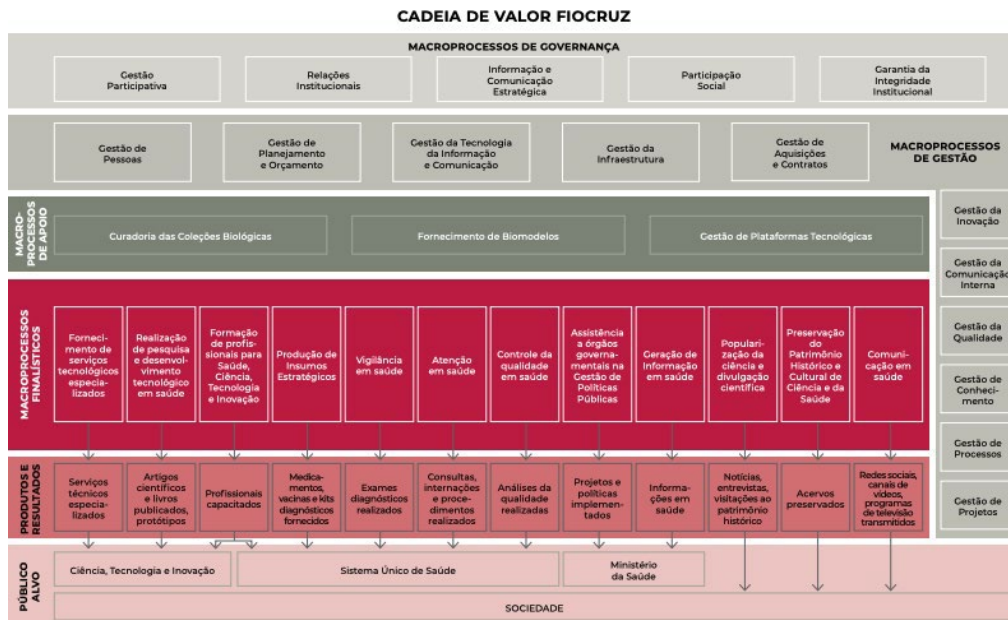


FIGURA 1. Cadeia de Valor Fiocruz

Fonte: Fundação Oswaldo Cruz (2022)

Um dos Macroprocessos estratégicos definidos pela Fiocruz, encontra-se a área de “Informação e Comunicação” que “busca contribuir para a democratização de informações e conhecimentos na área da saúde, além de dar suporte à formulação e implantação de políticas, programas e intervenções no setor” (Fundação Oswaldo Cruz, 2023). Dentro deste contexto, foi implementado em 2014 a **Política de Acesso Aberto ao Conhecimento** para disponibilizar o conteúdo integral da produção intelectual produzida, como também fortalecer os mecanismos de preservação da memória institucional. No Art.2º. da Política, ficou determinado que o **Arca – Repositório Institucional da Fiocruz** seria o principal instrumento de realização do Acesso Aberto, de forma a reunir, preservar e dar visibilidade à produção da Instituição para toda a sociedade. O Repositório passou a ser o local para o depósito de todas as tipologias e das coleções mandatórias (artigos, dissertações e teses dos programas de pós-graduação da Fiocruz). A governança do Repositório atua de modo articulado e colaborativo com a Rede de Bibliotecas da Fiocruz, tendo em

sua estrutura o Comitê da Regulamentação; o Comitê Gestor e os Núcleos de Ciência Aberta (NCA)<sup>1</sup>, da qual o Bibliotecário faz parte.

A preocupação com a preservação deste acervo digital passou a ser determinante para o seu funcionamento, tendo em vista a velocidade e o avanço das tecnologias da informação e da internet que criaram uma nova etapa para os periódicos científicos, inaugurando novas formulações de interoperabilidade e padrões de cooperação.

Os repositórios foram criados como estratégias (Via Verde) para viabilizar o acesso aberto aos documentos produzidos e disseminar a produção científica. O desenvolvimento dessas tecnologias produziu novos formatos e canais de comunicação expandindo assim, o fluxo da pesquisa científica, desde a sua produção até chegar aos usuários finais (Lanzellote, 2019).

Para estabelecer autenticidade, confiança e credibilidade no repositório digital é necessário o atendimento a determinados requisitos, padrões e critérios estabelecidos em normas que sistematizam o formato, etapas e resultados para uma gestão bem sucedida e confiável durante o processo de armazenagem, preservação dos metadados e dos objetos digitais depositados (OCLC/RLG, 2002). Holanda (2022) apresenta algumas competências e conhecimentos necessários para a realização do processo de preservação digital e auditoria que são fundamentais para os sistemas de gestão. Márdero Arellano (2019) destaca que

todo repositório confiável deve incluir atributos que sustentem os seguintes aspectos: a) responsabilidade administrativa; b) viabilidade organizacional; c) sustentabilidade financeira; d) adequabilidade tecnológica e procedimental; e)

---

<sup>1</sup> Os Núcleos de Ciência Aberta são "Instâncias colegiadas de caráter executivo, atuantes em todas as unidades da Fiocruz, responsáveis pela operação, participação, promoção e monitoramento das ações de implantação da Ciência Aberta, envolvendo ações para avaliar a viabilidade da disponibilização de dados e informações científicas, conforme disposto em regulamentação específica" (Fundação Oswaldo Cruz, 2021).

sistema de segurança; f) responsabilidade de procedimentos (certificação)". O repositório deve fornecer evidências para mostrar que opera um sistema de gerenciamento de conteúdos digitais e metadados adequado para garantir a integridade e a autenticidade durante os processos de armazenamento, arquivamento e acesso.

Nesse sentido, podemos elencar como um dos principais desafios enfrentados pelos Repositórios, às práticas de preservação digital, que precisam incluir em suas políticas institucionais, implementações tecnológicas, recursos humanos capacitados e conhecimento dos processos e critérios normativos fundamentadas em procedimentos técnicos, políticos e administrativos (Santos; Flores, 2017).

Na Fiocruz, o Repositório Arca que gerencia a produção técnico-científica se tornou uma rica fonte de informação e promoção para novos conhecimentos que estão alinhados com o Movimento de Acesso Aberto, através do estabelecimento da sua Política de Acesso Aberto. A importância da gestão desse conteúdo intelectual atende a missão da Fiocruz que tem como propósito

Produzir, disseminar e compartilhar conhecimentos e tecnologias voltados para o fortalecimento e a consolidação do Sistema Único de Saúde (SUS) e que contribuam para a promoção da saúde e da qualidade de vida da população brasileira, para a redução das desigualdades sociais e para a dinâmica nacional de inovação, tendo a defesa do direito à saúde e da cidadania ampla como valores centrais (Fundação Oswaldo Cruz, 2023).

Como a questão da Preservação se tornou fundamental para os acervos científicos da Fiocruz, foi lançado em 2018, a **Política de Preservação dos Acervos Científicos e Culturais**, que inclui ações de adequação a padrões e requisitos internacionais, e que foram debatidos por um Grupo de Trabalho, do qual participa a equipe executiva do Repositório Arca. Esse trabalho está alinhado com as premissas da Preservação Digital e Certificação que objetivam respeitar e manter as características originais dos objetos digitais e dos

metadados, respeitando os critérios e diretrizes estabelecidos nas normas internacionais para repositórios confiáveis. Segundo Thomaz (2007) “a certificação tornou-se um componente-chave para repositórios digitais contemporâneos” com o objetivo de avaliar e medir os critérios e requisitos, visando assim, definir ao longo do tempo, a autenticidade e a confiabilidade do processo.

A preservação digital é um processo que deve ser efetivado para garantir que o acesso aos objetos digitais armazenados ocorra de forma permanente, e a forma encontrada pela comunidade acadêmica para firmar a confiança em um repositório digital é através do atendimento a determinados padrões e critérios que, quando devidamente documentados, indicam evidência para “oferecer acesso confiável e de longo prazo aos recursos digitais por ele gerenciados para sua comunidade-alvo, agora e no futuro” (RLG/OCLC, 2002).

Várias iniciativas da área de preservação digital buscam sistematizar as evidências de que um repositório digital é confiável, e entre as mais reconhecidas estão a Trustworthy Repositories Audit and Certification: Criteria and Checklist (mais conhecida como TRAC), desenvolvida em 2007 pela OCLC/RLG e pela força-tarefa da Administração Nacional de Arquivos e Registros dos Estados Unidos (NARA). Este documento serviu de base para a norma ISO 16363:2012 (Audit and Certification of Trustworthy Digital Repositories), que divide seus requisitos em três categorias: Infraestrutura Organizacional, Gestão dos Objetos Digitais, e Gestão de Risco de Infraestrutura e Segurança.

Diante desse cenário, em 2019, a equipe executiva do Repositório Arca e uma Bibliotecária do Instituto de Tecnologia de Fármacos da Fiocruz (Farmanguinhos) deram início ao processo de autoavaliação do Repositório com relação aos critérios referentes à primeira categoria - de Infraestrutura Organizacional. O resultado do diagnóstico mostrou que o Arca atendia totalmente a 12 requisitos (44%), atendia parcialmente a 5 (19%) e não atendia a 10 requisitos (37%). Lanzelloti (2019) afirmou que “tendo em vista o objetivo



do estudo para avaliação do Repositório Arca, a revisão da literatura, apoiada no referencial teórico, foi essencial na construção do checklist, se tornando o instrumento para obter o diagnóstico final”. A análise dos critérios de confiabilidade definidos pela norma ISO 16363:2012 possibilitaram a verificação dos principais aspectos sobre o repositório que evidenciaram o nível de confiabilidade e segurança. É importante lembrar que a autoavaliação não é uma auditoria, porém sua realização é parte importante do processo de atualização e fortalecimento dos repositórios, incluindo a sua estratégia de preservação digital (Houghton, 2015).

Diversos autores (Corrado; Sandy, 2017; Houghton, 2015; Márdero Arellano; Oliveira, 2016) relatam que uma das vantagens da autoavaliação, assim como da auditoria de repositórios, “é não somente poder transmitir confiança para seus usuários, mas também identificar os pontos fortes e fracos e as oportunidades de melhoria”. Portanto, a autoavaliação ajudou a equipe do Repositório Arca a identificar as informações que não estavam documentadas ou que ainda não haviam sido formalizadas institucionalmente, destacando a importância da criação e manutenção de uma documentação formal sobre as práticas e procedimentos implementados (Lanzellotti, 2019).

O processo de autoavaliação realizado em 2019 foi muito importante, pois ofereceu subsídios para a elaboração do [Plano de Preservação Digital do Repositório Arca](#), trazendo uma análise das características, integridade e acessibilidade dos documentos. A fim de atender os requisitos de preservação para a gestão do repositório, foram realizados testes e apresentadas estratégias para atender a geração de SIP e exportação de pacotes de armazenamento de informação (AIP). Em 2020, foi lançado o Plano de

Preservação Digital<sup>2</sup>, que sistematizou e documentou o conjunto de orientações e procedimentos técnicos necessários para atender os requisitos mínimos para preservação dos objetos digitais armazenados.

Tendo em vista a necessidade de retomar essa atividade, a equipe executiva do Repositório Arca submeteu um Projeto em 2021 para o PIPDT/ICICT/FIOCRUZ 2021/2023<sup>3</sup>, intitulado “Autoavaliação de Confiabilidade do Repositório Institucional Arca pela norma ISO-16363:2012: avaliação dos resultados”, e contou com a participação de um Arquivista do Arquivo Nacional e de dois profissionais Bibliotecários de duas outras Unidades da Fiocruz – Biblioteca de Manguinhos e Biblioteca do Instituto Fernandes Figueira. Para o projeto foi definido a necessidade de fazer uma atualização nas perguntas utilizadas em 2019, com a inclusão de novos critérios descritos nas três categorias: Infraestrutura Organizacional, Gestão dos Objetos Digitais e Gestão de Risco de Infraestrutura e Segurança.

### **Justificativa**

Como o processo de autoavaliação do Repositório Arca já tinha sido realizado há mais de dois anos, foi preciso estabelecer uma nova estratégia para fornecer subsídios visando a revisão e o aperfeiçoamento do Plano de Preservação Digital e do processo de certificação. Por esse motivo, optou-se pela submissão do projeto ao PIPDT/Fiocruz para conseguir verbas para a execução do trabalho.

---

<sup>2</sup> A implementação do Archivematica não foi concluída no referido ano, tendo em vista o pedido da área de informática para concluir o processo de migração do sistema, que encontrava-se na versão 4.7 para a nova versão do DSpace.

<sup>3</sup> “O PIPDT é um Programa de Indução à Pesquisa e ao Desenvolvimento Tecnológico, sendo uma oportunidade para que os profissionais do Icient, Unidade da Fiocruz, para que possam concretizar projetos em consonância com a missão do instituto. Desde 2003, o programa tem fomentado a produção científica e tecnológica nos campos da Informação e da Comunicação em Saúde” (Fundação Oswaldo Cruz, 2023). Disponível em: <<https://pipdt.icient.fiocruz.br/>>

## Objetivo Geral

Realizar uma nova autoavaliação do Repositório Arca em relação aos requisitos de confiabilidade recomendados pela norma ISO 16363:2012 e compará-las com os resultados obtidos em 2019.

## Objetivos Específicos

- a) Realizar um levantamento bibliográfico sobre o tema, e a partir das leituras dos materiais coletados, definir os os requisitos determinados para estabelecimento de um repositório digital confiável;
- b) Fornecer subsídios para a revisão e aperfeiçoamento do Plano de Preservação Digital do Repositório Arca;
- c) Replicar o conhecimento adquirido no processo de autoavaliação para os Núcleos de Ciência Aberta (NCAs) da Fiocruz;
- d) Disseminar o conhecimento obtido para os membros da Rede Sudeste de Repositórios Digitais e para outras instituições que desejem realizar a autoavaliação de seus repositórios.

## Metodologia

Para contextualizar este trabalho é importante fazer um breve histórico do percurso da autoavaliação realizada em 2019, e que resultou no diagnóstico descrito na dissertação de mestrado intitulado **“Avaliação de confiabilidade do Repositório Arca da Fiocruz”** de autoria da Bibliotecária Michelle Frazão Lanzelloti<sup>9</sup>. O processo foi baseado nos critérios de confiabilidade contidos nas diretrizes apresentadas pelo CONARQ e pela norma ISO 16363:2012, que utilizam o modelo de referência OAIS para verificação de um conjunto de itens nas categorias Infraestrutura Organizacional, Gestão de Objetos Digitais e Infraestrutura Tecnológica e de Segurança. A partir do estudo desses documentos, foi necessário examinar a documentação disponível, os termos de direitos e o acesso ao sistema.

Para ilustrar o procedimento realizado em 2019, informamos que foi elaborado um checklist concluído com 27 perguntas, sendo o Repositório avaliado na sua totalidade incluindo as 26 comunidades e as 23 coleções, mensurada com a escala de Likert com os seguintes resultados, demonstrados no Quadro 1:

<b>QUADRO 1 – SÍNTESE DO RESULTADO DA ANÁLISE DE CONFIABILIDADE DO ARCA, REALIZADO EM 2019 CATEGORIAS</b>	<b>GRAU 0</b>	<b>GRAU 1</b>	<b>GRAU 2</b>	<b>GRAU 3</b>	<b>GRAU 4</b>
Infraestrutura Organizacional	3 (37 %)	-	-	-	5 (63 %)
Gestão de Objetos Digitais	3 (27 %)	2 (18 %)	1 (9 %)	2 (18 %)	3 (27 %)
Infraestrutura Tecnológica e de Segurança	4 (50 %)	-	-	-	4 (50 %)

Resumo:

- a)** Atendia totalmente a 12 requisitos da Norma (44%) – Grau 4;
- b)** Atendia parcialmente a 5 requisitos (19%) – Grau 1 a Grau 3;
- c)** E não atendia a 10 requisitos (37%) – Grau 0.

De acordo com o “levantamento feito, analisado, justificado e avaliado através do somatório de todos os graus e divididos pelo total máximo, considerando o total máximo como se o repositório tivesse atingido grau 4 em todos os quesitos, o Repositório Arca transmite ou garante a confiança de 53,7%” (Lanzellote, 2019).

### **Nova autoavaliação 2021/2023**

O projeto iniciado em 2021, teve como primeira etapa fazer um levantamento bibliográfico sobre o tema com a descrição dos requisitos de confiabilidade recomendados pela norma ISO 16363:2012. Após a finalização do levantamento das referências, a equipe passou para a fase de estudos e leitura. Para documentar o processo foram utilizados como evidências: a autoavaliação realizada em 2019; os documentos existentes que regem o Repositório; a [Política de Acesso Aberto ao Conhecimento](#); a [Política de Preservação de Acervos da Fiocruz](#); o [Plano Operativo e o Plano de Backup](#); os [Termos de Uso](#), o [Termo de Cessão de Direitos Autorais](#); o [Manual de preenchimento de metadados](#) e outros documentos internos.

Com os resultados obtidos na primeira etapa de autoavaliação do Repositório Arca, realizada em 2019, foi possível identificar os pontos atendidos e quais seriam as ações necessárias para resolver as recomendações pendentes.

Total Máximo 34 perguntas (34 x 4) = 136 pontos

28 perguntas x (Grau 4) 0 perguntas x (Grau 3) 3 perguntas x (Grau 2) 0 perguntas x (Grau 1) 2 perguntas x (Grau 0) 1 pergunta (Não aplicável)

Total do Arca = 118 pontos (86,7%)

A segunda etapa envolveu o processo de revisão das perguntas e respostas de acordo com os critérios da norma ISO 16.363:2012. Nessa parte, foi fundamental a participação do profissional do Arquivo Nacional, especializado em Certificação. Após esta fase, foi realizada a tabulação dos dados e a sistematização do conhecimento adquirido, através da apresentação dos resultados, de forma a contribuir para a construção e aperfeiçoamento do Repositório Arca.

## Resultados parciais e finais

A autoavaliação submetida em 2023, analisou antigos e novos critérios, a partir de um checklist revisado, concluído com 136 perguntas, incluindo as 26 comunidades e as 23 coleções, mensurada com a escala de Likert. Os resultados estão demonstrados no Quadro 2.

<b>QUADRO 2 – SÍNTESE DO RESULTADO DA ANÁLISE DE CONFIABILIDADE DO ARCA, REALIZADO EM 2023 CATEGORIAS</b>	<b>GRAU 0</b>	<b>GRAU 1</b>	<b>GRAU 2</b>	<b>GRAU 3</b>	<b>GRAU 4</b>	<b>NÃO APLICÁVEL</b>
Infraestrutura Organizacional	2 (6 %)	0 (0 %)	3 (9 %)	0 (0 %)	28 (82 %)	1 (3 %)
Gestão dos Objetos Digitais	10 (14 %)	20 (29 %)	10 (14 %)	2 (3 %)	28 (40 %)	-
Gestão de Risco de Infraestrutura e Segurança	-	-	-	-	-	-

(Pereira et al., 2023 apud Lanzellote, 2019)

Resumo:

- a)** Atende totalmente a 56 requisitos da Norma (41,2%) – Grau 4;
- b)** Atende parcialmente a 35 requisitos (26%) – Grau 1 a Grau 3;
- c)** Não atende a 12 requisitos (8,8%) – Grau 0;
- d)** Não aplicável – 1 requisito.

Em relação aos quesitos da categoria “Infraestrutura Organizacional”, o Arca garantiu 118 pontos. Dentre as 34 perguntas que correspondem ao total máximo de 136 pontos, o repositório conseguiu 86,7% da nota máxima nessa categoria, demonstrado na Figura 2.

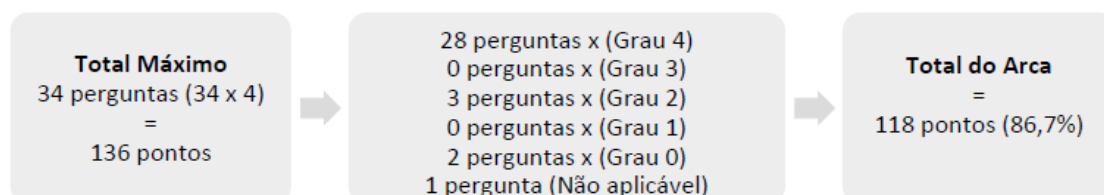


FIGURA 2. Infraestrutura Organizacional

Fonte: Pereira et al., 2023 apud Lanzellote, 2019

Referente aos quesitos da categoria “Gestão dos Objetos Digitais”, o Arca garantiu 158 pontos. Dentre as 70 perguntas que correspondem ao total máximo de 280 pontos, o repositório conseguiu 56,4% da nota máxima nessa categoria, conforme demonstrado na Figura 3.

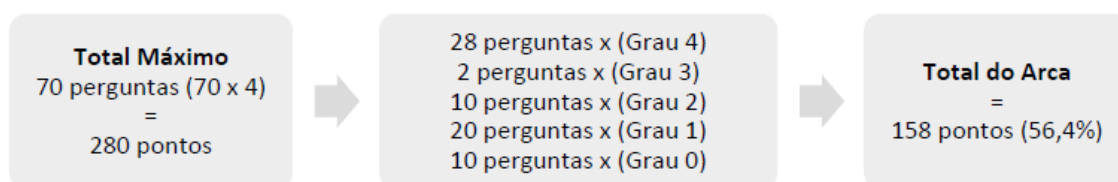


FIGURA 3. Gestão dos Objetos Digitais

Fonte: Pereira et al., 2023 apud Lanzellote, 2019

De acordo com o levantamento feito, analisado, justificado e avaliado através do somatório de todos os graus e divididos pelo total máximo, considerando o total máximo como se o repositório tivesse atingido grau 4 em

todos os quesitos, o Repositório Arca transmite ou garante a confiança de 66,3 %.

O comparativo entre as autoavaliações realizadas em 2019 e em 2023, que envolveu o processo de revisão das perguntas e respostas de acordo com os critérios da norma ISO 16.363:2012, está ilustrado no Gráfico 1 abaixo.

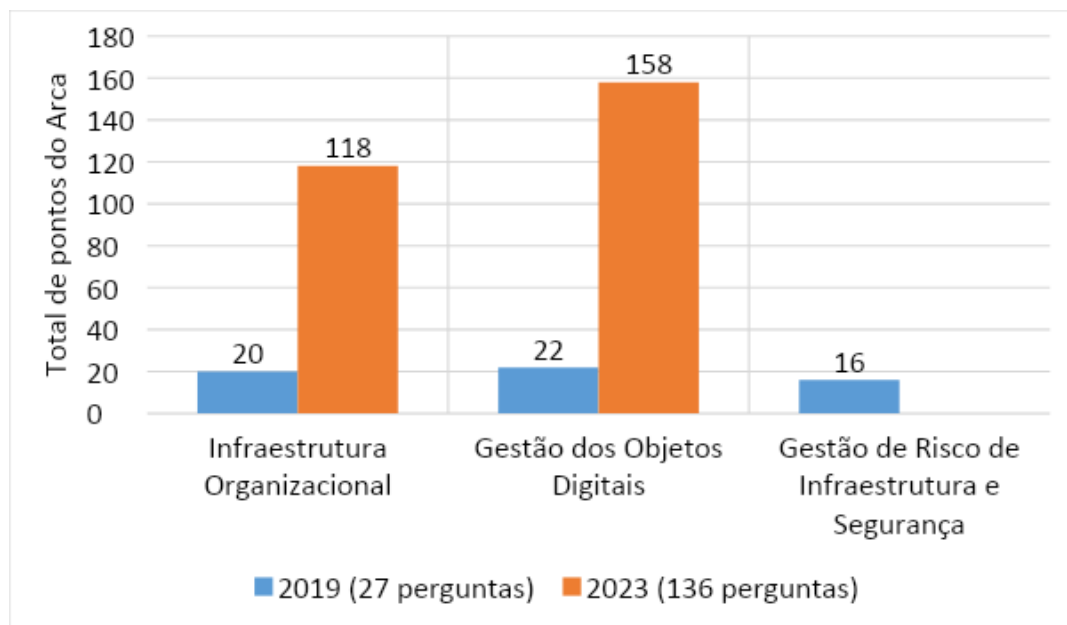


GRÁFICO 1. Comparativo da autoavaliação feita em 2019 e em 2023

Fonte: Pereira et al, 2023

### Considerações finais

A análise dos critérios de confiabilidade definidos pela ISO 16363:2012 e analisados nessa pesquisa representam os requisitos para que um repositório digital possa garantir sua confiabilidade e segurança. A autoavaliação ajudou a identificar e documentar dados sobre o repositório e fornecer informações para garantir autenticidade, confidencialidade e principalmente obter informações para que a preservação digital fosse realizada. Com isso, destacamos a importância da manutenção deste trabalho e o aprimoramento das práticas e procedimentos implementados de forma periódica. É importante lembrar que a



autoavaliação é parte importante do processo de atualização e fortalecimento dos repositórios, incluindo a estratégia de preservação digital. A produção de documentação referente ao processo vai poder servir como material para a avaliação de outros repositórios, usando como referência a Rede Sudeste de Repositórios Digitais para replicar o conhecimento adquirido. É válido informar ainda, que os profissionais que trabalharam no projeto adquiriram novas competências e habilidades na área de Preservação e Certificação.

Não tivemos como colocar todas as tabelas nesse trabalho pela extensão do número de páginas que excederia o total máximo permitido, mas pretendemos depositá-las no Arca – Repositório Institucional da Fiocruz<sup>4</sup> e no [Arca Dados – Repositório Institucional de Dados de Pesquisa da Fiocruz](#), porque entendemos que o resultado obtido no formato de dados, deveria estar num local para acesso e disseminação.

Concluindo, destacamos que as instituições que utilizam repositórios precisam entender que a autoavaliação deve ser um processo contínuo e recorrente para medir a confiabilidade dos sistemas. Nossa expectativa é que a certificação dos Repositórios se torne uma realidade no Brasil, e que possa contribuir para a melhoria da qualidade na produção intelectual de uma Instituição.

## Bibliografia

CONSELHO NACIONAL DE ARQUIVOS (2015). Diretrizes para a implementação de repositórios arquivísticos digitais confiáveis – RDC-Arq. <[http://www.conarq.gov.br/images/publicacoes\\_textos/diretrizes\\_rdc\\_arq.pdf](http://www.conarq.gov.br/images/publicacoes_textos/diretrizes_rdc_arq.pdf)>

---

<sup>4</sup> Disponível em: <<https://www.arca.fiocruz.br/handle/icict/58796>>. Temporariamente em acesso fechado. O acesso ficará aberto após a apresentação deste trabalho.

- CORRADO, E. M.; SANDY, H. M. (2017). *Digital preservation for libraries, archives and museums*. 2. ed.
- FUNDAÇÃO OSWALDO CRUZ. CASA DE OSWALDO CRUZ (2018). Política de preservação dos acervos científicos e culturais da Fiocruz. 72 p.
- FUNDAÇÃO OSWALDO CRUZ. PRESIDÊNCIA. COORDENAÇÃO GERAL DE PLANEJAMENTO ESTRATÉGICO (2022). Relatório de Gestão 2022. 139 p.
- HOLANDA, A. P. DE. (2022). Curso certificação de repositórios digitais. Rio de Janeiro: RBRD - Rede SudesteFiocruz/Icict; Fiocruz/Icict. 78 p. <<https://www.arca.fiocruz.br/handle/icict/55232>>
- HOUGHTON, B. (2015). Trustworthiness: Self-assessment of an Institutional Repository against ISO 16363-2012. *D-Lib Magazine*, 21, 3/4,1-5.
- MÁRDERO ARELLANO, M. Á. (2019). Preservação de acervos digitais em repositórios institucionais. In: Encontro da Rede Sudeste de Repositórios Institucionais, 1., 2019, Rio de Janeiro. Anais... Rio de Janeiro: Fiocruz/Icict/UFRJ. 77 p. <<https://www.arca.fiocruz.br/handle/icict/33333>>
- PEREIRA, C. B. M., QUEIROZ, C. F. DE, HOLANDA, A. P. DE, ARAUJO, L. D. DE, RODRIGUES, R. B., SILVA, A. A. DA, ... FERREIRA, T. M. DA C. (2023). Síntese de dados: confiabilidade do Arca - Repositório Institucional da Fiocruz conforme critérios da ISO 16363:2012. Rio de Janeiro: Icict. 45 p. <<https://www.arca.fiocruz.br/handle/icict/58796>>
- RLG/OCLC (2002). *Trusted digital repositories: attributes and responsibilities*. Mountain View, CA.: Research Library Group/OCLC.
- SANTOS, H. M.; FLORES, D. (2017). Os impactos da obsolescência tecnológica frente à preservação de documentos digitais. *Brazilian Journal of Information Science: Research Trends*, 11, 2. <<http://200.145.171.5/revistas/index.php/bjis/article/view/5550>>

# Formación de competencias para la Ciencia Abierta en la Universidad del Rosario (Colombia)

## **JULY ANDREA GARCÍA AMÉZQUITA**

CRAI Universidad del Rosario

[july.garcia@urosario.edu.co](mailto:july.garcia@urosario.edu.co)

## **MALGORZATA LISOWSKA NAVARRO**

CRAI Universidad del Rosario

[margarita.lisowska@urosario.edu.co](mailto:margarita.lisowska@urosario.edu.co)

### **EJE TEMÁTICO**

Comunicación académica, científica y cultural en abierto

### **RESUMEN**

La Universidad del Rosario en Colombia ha creado un Modelo Institucional de Ciencia Abierta y parte de la estrategia para la materialización lleva a cabo un programa de formación de competencias digitales llamado DigiSkills. Este programa busca promover la adopción de prácticas y recursos abiertos en la comunidad académica, y ofrece cursos y talleres en áreas relacionadas con la comunicación académica en abierto, datos de investigación en abierto, métricas de nueva generación, integridad científica y ciencia ciudadana. Este programa ha llegado a más de 3000 personas en sus dos años de implementación. El objetivo de la presente ponencia que se enmarca en el eje de Comunicación académica, científica y cultural en abierto, es presentar el programa de formación, sus componentes y los resultados obtenidos a la fecha.

### **PALABRAS CLAVES**

Ciencia abierta; competencias digitales; datos de investigación; integridad científica; cambio cultural; democratización del conocimiento. Open Science; cultural change; digital skills; research data; scientific integrity; democratization of knowledge.

## Introducción

El interés por la cultura de lo abierto en el mundo académico se remonta a las últimas décadas del siglo XX y se consolida en los primeros años de este siglo con las tres declaraciones internacionales sobre acceso abierto (las 3 B, Berlín, Bethesda y Budapest). Más recientemente, ha crecido el interés internacional por transparentar y democratizar la ciencia, por lo cual múltiples organismos, como la OCDE, Unión Europea o UNESCO han incorporado dentro de sus agendas el término de Ciencia Abierta.

En noviembre del año 2021 y luego de una amplia consulta mundial, los representantes de los gobiernos aprobaron la [Recomendación de UNESCO sobre Ciencia Abierta](#), que define la ciencia abierta como:

Un constructo inclusivo que combina diversos movimientos y prácticas con el fin de que los conocimientos científicos multilingües estén abiertamente disponibles y sean accesibles para todos, así como reutilizables por todos, se incrementen las colaboraciones científicas y el intercambio de información en beneficio de la ciencia y la sociedad, y se abran los procesos de creación, evaluación y comunicación de los conocimientos científicos a los agentes sociales más allá de la comunidad científica tradicional (UNESCO, 2021, p. 7).

Esta definición sin duda se constituye en un importante desafío para todos los actores involucrados, empezando por la sensibilización y capacitación sobre el tema a la comunidad académica. Es tanto así, que dentro de los principales objetivos y ámbitos de acción de las ya citadas recomendaciones de la UNESCO figura el de “invertir en recursos humanos, formación, educación, alfabetización digital y desarrollo de capacidades para la ciencia abierta” (UNESCO, 2021, p. 26). Es por eso que las instituciones de educación superior deben incursionar no solo en la creación de las políticas o lineamientos sobre Ciencia Abierta, sino también crear programas de formación a su comunidad.

Esta ponencia ilustra la experiencia de la Universidad del Rosario de Colombia, en la creación de un programa de formación de las competencias digitales, enfocados en Ciencia Abierta dirigido a diferentes públicos

académicos, ofrecido por el Centro de Recursos para el Aprendizaje y la Investigación - CRAI. Se mostrará la evolución del modelo de formación, el soporte conceptual, algunos ejemplos prácticos y los resultados obtenidos a la fecha.

### **Cultura de lo abierto en la Universidad del Rosario (Colombia)**

Para la Universidad del Rosario, promover la ciencia abierta es parte de su compromiso con el desarrollo de investigación de alta calidad, con un enfoque inclusivo, colaborativo y transparente que genere un impacto significativo y relevante para la sociedad. En este escenario, se ha avanzado en acciones concretas que apuntan a la apropiación y sensibilización de la comunidad académica Rosarista, sobre la cultura de lo abierto.

La reflexión institucional y el esfuerzo sostenido con el movimiento internacional de acceso abierto, uno de los pilares de la Ciencia Abierta, comenzó hace más de una década, con la participación activa en iniciativas, nacionales o regionales, como la Biblioteca Digital Colombiana, LA Referencia y la conferencia BIREDIAL-ISTEC, entre otras. Igualmente, la Universidad ha implementado las herramientas de materialización de acceso abierto, como el repositorio institucional E-docUR (Knudsen Quevedo & Nieto Arango, 2008) y el portal de revistas de acceso abierto (en 2010), además de la expedición, en 2017 —como primera universidad en Colombia—, de la Política Institucional de Acceso Abierto (Rosario, 2016) y la firma de la Declaración de Berlín. Estas acciones han sido premiadas con dos reconocimientos significativos: el Premio al Mejor Repositorio Institucional de Colombia y el Premio Nacional de Ciencia Abierta, en su categoría Acceso Abierto, otorgado por Colciencias (ahora, Ministerio de Ciencia, Tecnología e Innovación), en 2017.



GRÁFICA 1. Modelo de Ciencia Abierta Universidad del Rosario - Pilares

A partir de esta exitosa experiencia y concordante con las tendencias internacionales, la Universidad incorporó dentro de sus apuestas, la creación de una estrategia de Ciencia Abierta articulada en ocho pilares: comunicación académica en abierto; datos de investigación en abierto; métricas de nueva generación; integridad científica; ciencia ciudadana; infraestructura y recursos; habilidades y educación; y reconocimientos por publicar en abierto.

Este modelo fue validado por la comunidad académica representada por más de 70 miembros, en mesas de trabajo presenciales y virtuales, y permitieron además reconocer las prioridades, intereses y expectativas de los investigadores y de otros actores de la comunidad sobre Ciencia Abierta. El documento final de los lineamientos institucionales de Ciencia Abierta, se puede consultar en el [Portal Institucional de Ciencia Abierta](#).

Igualmente se establecieron diferentes objetivos, transversales a la toda la comunidad académica de la Universidad del Rosario, independiente de las diferencias disciplinares, dentro de los que se encuentran:

- Sensibilizar e incentivar la adopción de los lineamientos de Ciencia Abierta en toda la comunidad académica Rosarista y con los actores externos que participan de alguna forma en los procesos de investigación.
- Propender por la inserción de la Ciencia Abierta en los currículos de los programas académicos con el fin de contribuir a una educación abierta y romper posibles fragmentaciones en la gestión del conocimiento.
- Desarrollar capacidades en la comunidad Rosarista para la creación, acceso, uso, reutilización, rediseño y comunicación de Recursos Educativos Abiertos.
- Promover el almacenamiento de preprints y artículos científicos en acceso abierto en los casos en que las políticas editoriales de las revistas lo permitan, y el almacenamiento de otro tipo de recursos científicos y educativos; así como una amplia divulgación de los beneficios de las plataformas abiertas institucionales.
- Garantizar que el 100% de los proyectos de investigación financiados con recursos institucionales desarrollen el plan de gestión de datos con el fin de garantizar la homogenización de sus prácticas; así como una amplia sensibilización sobre los recursos y servicios de apoyo con los que cuentan los investigadores para la gestión de datos en abierto.
- Diseñar y ofertar cursos de formación sobre las métricas de nueva generación y sus impactos.
- Ofertar servicios de capacitación y acompañamiento para garantizar investigaciones e innovaciones justas y responsables; prácticas académicas en el marco de la transparencia y la integridad; y una mayor cantidad de publicaciones en abierto.
- Implementar acciones que faciliten la participación de la sociedad en la formulación, desarrollo de la investigación, así como en la socialización y retroalimentación de los resultados.

## Formación de competencias

La materialización de esta ambiciosa apuesta representa un importante cambio cultural, en donde lo abierto se convierte en la nueva normalidad y todas las partes que intervienen en los procesos académicos y de investigación, adopten los principios, las políticas y las prácticas de la Ciencia Abierta.

Para apoyar esta estrategia institucional y la transición a la cultura de lo abierto, el Centro de Recursos para el Aprendizaje y la Investigación CRAI de la Universidad del Rosario reorientó su formación de competencias informacionales hacia un modelo de formación de competencias digitales llamado DigiSkills<sup>1</sup>, que apoya la adopción de enfoques, prácticas y recursos abiertos por parte de la comunidad rosarista.

DigiSkills fue creada en 2019<sup>2</sup> se basa en cinco líneas de formación i) información y datos; ii) creación de contenidos; iii) seguridad y conciencia digital; iv) solución de problemas en entornos digitales; y v) comunicación y visibilidad y tiene dentro de sus aspiraciones, relacionadas con la cultura de lo abierto, las siguientes:

- Aportar a un cambio de cultura, donde lo abierto sea la nueva normalidad y los Rosaristas se distingan por ser prosumidores éticos que usan contenidos en abierto y aportan a la democratización del conocimiento.

---

<sup>1</sup> DigiSkills define la competencia digital como el “conjunto de conocimientos, habilidades y actitudes que permiten el aprovechamiento del ecosistema digital abierto, a través de su uso creativo, crítico y seguro; para participar de forma significativa en la sociedad y desarrollar un aprendizaje a lo largo de la vida”.

<sup>2</sup> Para la creación del Modelo DigiSkills se tomó como referente el Marco de Competencias Digitales para la Ciudadanía de la Unión Europea <<https://somos-digital.org/digcomp/>>



- Fortalecer las competencias para la ciencia abierta con la idea de que un futuro cercano, esta sea la forma predeterminada de educar, investigar y comunicar la ciencia.
- Acreditar las competencias desarrolladas por los participantes del programa de formación, mediante insignias digitales que, además, incentiven a otros miembros de la comunidad académica a transformar sus prácticas hacia la cultura de lo abierto.

De esta forma, las cinco líneas de formación de competencias digitales se alinearon con los cinco pilares generales del modelo institucionales de Ciencia Abierta, generando el siguiente marco de formación:

<b>Comunicación académica en abierto</b>					
<b>Datos de investigación en abierto</b>					
<b>Métricas de nueva generación</b>					
<b>Integridad Científica</b>					
<b>Ciencia ciudadana</b>					
	<b>Información y datos</b>	<b>Creación de contenidos</b>	<b>Seguridad y conciencia digital</b>	<b>Solución de problemas en entornos digitales</b>	<b>Comunicación y visibilidad</b>

GRÁFICO 2. Matriz de Formación de competencias digitales para la Ciencia Abierta UR

Guiado por este marco de acción, se diseñó el siguiente portafolio de talleres, capacitaciones y asesorías que fortalezcan las competencias sobre Ciencia Abierta en la comunidad académica:



GRÁFICO 3. Formación de competencias digitales para la Ciencia Abierta UR – Portafolio

El reto del pilar **Comunicación académica en abierto**, es el de “Consolidar la cultura de publicar en abierto, desarrollar lineamientos sobre recursos educativos abiertos (REA) y generar nuevas formas de publicación”, por lo cual se ofrecen cursos, talleres y asesorías para:

- Búsqueda y recuperación de fuentes de información, recursos multimedia y datos abiertos
- Aprovechamiento de recursos abiertos
- Evaluación de revistas en acceso abierto
- Uso de software y tecnologías en abierto para la creación de contenidos como mapas de conocimiento, materiales didácticos, casos de estudio, imagen, audio, video
- Reutilización de REA
- REA como estrategia de aprendizaje

- Propiedad intelectual en entornos digitales
- Gestión de contraseñas y permisos en plataformas colaborativas
- Depósito de Recursos Educativos Abiertos (Multimedia, aplicaciones informáticas, contenido didáctico) en el Repositorio Institucional
- Depósito de preprints, literatura gris, en el Repositorio Institucional
- Capacitación en Open Journal Systems
- Comunicación de resultados de investigación a través de redes sociales
- Opciones de publicación en abierto (vía verde, dorada, híbrida)
- Estrategias de publicación en abierto (revistas, contratos, relaciones con editores, modelos de financiación)

En el pilar Datos de investigación en abierto, es importante mencionar que la Universidad del Rosario cuenta con una estrategia institucional para la gestión de datos de investigación, constituida por tres componentes principales: i) **Política institucional para la gestión de datos de investigación** (2019), que define lineamientos generales para el proceso de gestión de datos de investigación. ii) un Repositorio de Datos de Investigación, espacio perfecto para almacenar, gestionar y difundir los datos, que permite centralizar su acceso, y compartirlos a la comunidad, aumentando su impacto y visibilidad. iii) Servicios de apoyo en las diferentes etapas del ciclo de vida de los datos de investigación, tales como:

- Taller de sensibilización: ¿Qué son los datos abiertos?
- Asesorías personalizadas para la construcción del plan de gestión de datos
- Análisis y visualización de datos
- Minería de datos y texto
- Limpieza de datos
- Reproducibilidad y reutilización de datos
- Metadatos para colecciones y conjuntos de datos digitales

- Almacenar, guardar, archivar y conservar datos
- Políticas institucionales, nacionales e internacionales sobre datos de investigación
- Principios DATA FAIR
- Software para procesamiento de datos
- Conservación y publicación de datos

En el pilar Métricas de nueva generación, uno de los desafíos es “ampliar la gama de bibliometría para cubrir nuevas formas de producción, como datos de investigación y software de investigación, con nuevas métricas; y también acordar principios para el uso responsable de las métricas” (League Of European Research Universities, s. f., p. 18). Es por esto que se ofrecen cursos, talleres y asesorías en:

- Impacto de la investigación bibliometría, almetrics, y métricas de nueva generación
- Monitoreo de uso de contenidos multimedia en plataformas libres
- Estrategias de visibilidad web para autores y sus productos académicos
- Visualización y análisis de estadísticas de RRSS
- Visibilidad de las fortalezas individuales e institucionales HUB UR

En el pilar Integridad Científica el principal reto es “Ofrecer al investigador y al estudiante de la Universidad del Rosario mecanismos prácticos para aplicar la Ciencia Abierta con responsabilidad y transparencia”. Es por esto que se ofrecen cursos, talleres y asesorías en:

- Licencias abiertas y dominio público
- Derechos de autor y propiedad intelectual en el entorno digital
- Prácticas éticas en la investigación
- Uso ético de contenidos digitales (Citación y referenciación)

- Licenciamiento en el entorno digital
- Investigación e innovación responsable
- Ética de las publicaciones

En el pilar Ciencia Ciudadana, el principal reto es Ofrecer a los investigadores un conjunto de herramientas para que puedan entender fácilmente la ciencia abierta y aplicarla en sus procesos de investigación. Es por esto que se ofrecen cursos, talleres y asesorías en:

- Creación de experiencias participativas en plataformas abiertas (RRSS, wikis, transmedias, blogs)
- Creación de encuestas y materiales documentales
- Talleres y capacitaciones de creación de contenidos para diferentes públicos
- Licencias y derechos de autor en la cocreación
- Privacidad y anonimización de datos
- Uso de plataformas colaborativas
- Marketing científico
- Gestión de eventos para interacciones en línea crowdsourcing

De forma paralela, el CRAI UR ha insertado la formación de competencias digitales en el currículo, a través de las asignaturas para pregrado como “Más que Google, herramientas para el aprovechamiento de la transformación Digital” (1 crédito); “Competencias digitales para abogados” (2 créditos), “Competencias digitales para profesionales en Ciencias de la Salud” (2 créditos); y el curso para especializaciones “Información académica: Busca, Cita y Comunica”. En todas estas asignaturas, se incluyen aspectos de sensibilización en coherencia con el Modelo Institucional de Ciencia Abierta, incluyendo actividades que implican la promoción del acceso abierto; la

integridad académica; y la búsqueda y recuperación de datos, fuentes y contenidos multimedia en abierto.

## Resultados

La estrategia de formación de competencias digitales para Ciencia Abierta, guiada por un Centro de Recursos para el Aprendizaje y la Investigación CRAI, es exitosa por la alta posibilidad de llegar a los investigadores de forma personalizada, así como de insertarse en el currículo a través de asignaturas que cubren todos los niveles de formación.

La utilización del portafolio de cursos, talleres y asesorías ofertados mediante el programa de formación de competencias para la Ciencia Abierta, ha tenido una aceptación cada vez más importante, llegando entre enero de 2022 y mayo de 2023 a 3.189 usuarios de una comunidad académica con 14.000 miembros activos.

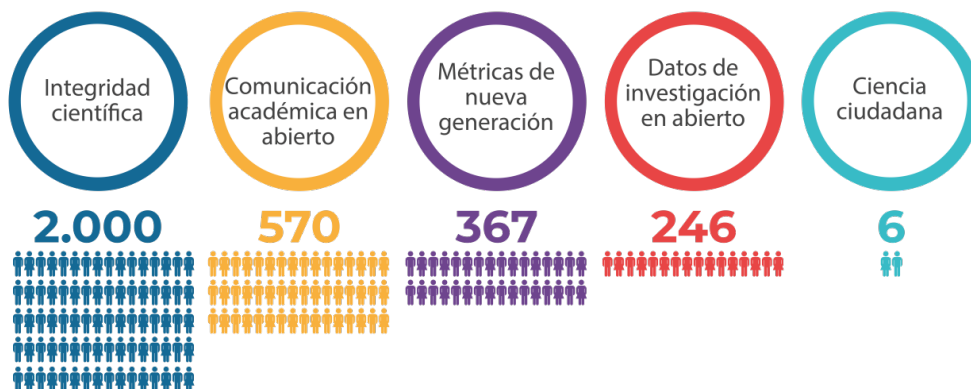


GRÁFICO 4. Participantes del Programa de Formación de Competencias para la Ciencia Abierta 2022 – mayo 2023

Como se puede apreciar en la gráfica, dentro de los pilares con mayor cobertura en términos de formación, se encuentran el de Integridad científica, lo cual puede explicarse por qué desde 2017 la Universidad ha alineado sus

procesos académicos e investigativos hacia el fortalecimiento de buenas prácticas científicas como la citación y referenciación; la creación de productos en el marco de la Propiedad Intelectual; el licenciamiento y el uso ético de los contenidos digitales. Esta formación, en general se realiza a través de cursos, talleres y capacitaciones grupales.

En el caso de los pilares de Comunicación académica en abierto, Métricas de nueva generación y Datos de investigación en abierto, la mayoría de investigadores son asesorados de forma individual, por tanto, la cobertura es menor, pero con un alto impacto y visibilidad institucional. A manera de ejemplo, se resalta el apoyo para la creación del Plan de Gestión de Datos de Investigación, el cual es un procedimiento diseñado en conjunto con la Dlel, que aplica para todos los proyectos de investigación que reciban financiación institucional.

Por su parte, los servicios asociados al pilar de Ciencia ciudadana, tienen un menor tiempo de implementación y por tanto de apropiación por parte de la comunidad de investigadores, sin embargo, es posible apreciar algunos resultados del acompañamiento para la divulgación de proyectos de ciencia ciudadana en la serie “Voces de la comunidad” disponibles en [este enlace](#).

Por otra parte, es importante mencionar que para nutrir el portafolio de cursos, talleres y asesoría para la formación de competencias en Ciencia Abierta y garantizar una oferta innovadora y pertinente, es necesario conocer las particularidades de la comunidad, ya que las prácticas de investigación tienen diferencias disciplinares y el nivel de apropiación de los investigadores puede ser disímil.

Es por esto que, en el año 2020, una de las unidades académicas con mayor trayectoria y resultados de investigación, realizó un diagnóstico para conocer las capacidades y los retos que, en materia de Ciencia Abierta han venido acumulando. En este ejercicio participaron más de 200 profesores, encontrando que cerca del 44%, conocía los conceptos de Ciencia Abierta con

un mayor énfasis en algunos de ellos, como es el caso de la comunicación académica en abierto e integridad científica.

Este tipo de ejercicios, permite además identificar los aspectos que requieren mayores esfuerzos para lograr su fortalecimiento y su consolidación, así como para identificar, orientar y trazar las acciones necesarias y focalizarlas para avanzar hacia la apropiación de los lineamientos de Ciencia Abierta a nivel institucional.

### **Conclusiones**

Las bibliotecas o centros de recursos para el aprendizaje y la investigación pueden tener una importante participación en la materialización de la Ciencia Abierta en las instituciones, debido a su experiencia en la formación de usuarios, sobre todo las que han transitado en las han transitado a la formación de competencias digitales y cuentan con equipos multidisciplinares con competencias especializadas y experiencia en investigación.

Los programas de formación de competencias para la Ciencia Abierta requieren de un alto nivel de personalización y para lograr este propósito, es necesario realizar ejercicios diagnósticos por áreas disciplinares.

Las competencias digitales, son una de las competencias clave para el aprendizaje permanente, en una época donde la revolución digital y la sociedad de la información, requieren de ciudadanos hábiles, con una amplia conciencia digital y sentido crítico frente a las posibilidades de las nuevas tecnologías y los grandes volúmenes de información.

### **Bibliografía**

KNUDSEN QUEVEDO, H.-P., & NIETO ARANGO, L. E. (2008). Decreto Rectoral 1037: Por el cual se establece la política de creación, funcionamiento



y utilización del repositorio institucional.  
<<http://repository.urosario.edu.co/handle/10336/641>>

LEAGUE OF EUROPEAN RESEARCH UNIVERSITIES. (s. f.). Open Science and its role in universities: A roadmap for cultural change.  
<<https://www.leru.org/files/LERU-AP24-Open-Science-full-paper.pdf>>

ROSARIO, U. del. (2016). Política Institucional de Acceso Abierto a la producción científica, académica e institucional de la Universidad del Rosario. <<http://repository.urosario.edu.co/handle/10336/12300>>

UNESCO. (2021). Recomendación de la UNESCO sobre la Ciencia Abierta. UNESCO.  
<[https://unesdoc.unesco.org/ark:/48223/pf0000379949\\_spa](https://unesdoc.unesco.org/ark:/48223/pf0000379949_spa)>

UNIVERSIDAD DEL ROSARIO. (2019). Política institucional para la gestión de datos de investigación de la Universidad del Rosario.  
<<http://repository.urosario.edu.co/handle/10336/19340>>

# Acesso Aberto Verde: breves reflexões a partir de casos da produção intelectual publicada por Instituições de Ensino brasileiras

## **JOSIMARA DIAS BRUMATTI**

SUPERINTENDÊNCIA DE DOCUMENTAÇÃO DA UNIVERSIDADE FEDERAL FLUMINENSE

[josimaradias@id.uff.br](mailto:josimaradias@id.uff.br)

## **SIMONE DA ROCHA WEITZEL**

PROGRAMA DE PÓS-GRADUAÇÃO EM BIBLIOTECONOMIA (UNIRIO)

[sweitzel@unirio.br](mailto:sweitzel@unirio.br)

### **EIXO TEMÁTICO**

Comunicación académica, científica y cultural en abierto

### **RESUMO**

Apesar do sucesso do Acesso Aberto no Brasil, especialmente no Acesso Aberto Diamante, é preciso reconhecer que ainda há um déficit em seu desempenho em determinadas áreas do conhecimento que publicam em periódicos comerciais (Ortellado, 2008). Soma-se a isto o fato de que os pesquisadores brasileiros ainda têm preferência por publicarem em periódicos com alto fator de impacto, devido não somente a o prestígio maior, mas pelas condições exigidas para obter financiamentos e verbas de órgãos públicos de fomento, que avaliam dados quantitativos da produção científica (Furmival, 2015). Tais anacronismos podem afetar o desenvolvimento do AA no país. A pesquisa tem por objeto de estudo a produção científica publicada em periódicos científicos gerada pelas Instituições de Ensino Superior (IES) e depositada em Repositórios Institucionais (RI), visando caracterizar o Acesso Aberto Verde no Brasil. Observamos: a) apenas 3 % dos artigos estão depositados em RIs; b) cerca de 25 % da produção nacional apresentou barreiras de acesso; c) a pesar de 75 % da publicação nacional, estar em Acesso Aberto, identificamos que este número expressivo, foi alavancado pela adesão do AA Híbrido (23 %). Desta forma, constatamos que o Acesso Aberto está ligado à tradição e cultura de determinada área do conhecimento no fazer científico.

### **PALAVRAS CHAVE**

Acesso Aberto; Acesso Aberto Verde; Produção intelectual; Brasil. Open Access; Open Access Green; Intellectual production; Brazil.

A convergência de tecnologias com o advento da internet e as dificuldades enfrentadas pela comunidade científica para acessar e disseminar a informação científica despertou o surgimento de um movimento em prol do acesso e disseminação ampla e irrestrita da produção científica revisada por pares a toda comunidade científica do mundo beneficiando, sobretudo, a inclusão de pesquisadores do hemisfério sul e até mesmo de pessoas leigas interessadas no assunto.

Acesso aberto foi o termo adotado pela Budapest Open Access Initiative, mais conhecida por sua sigla BOAI (2002), que defende a “disponibilidade gratuita na internet, permitindo a qualquer usuário a ler, baixar, copiar, distribuir, imprimir, buscar ou usar essa literatura” (BOAI, 2012).

A literatura sobre o acesso aberto é unânime ao preconizar iniciativas anteriores que vão desde os primeiros periódicos eletrônicos do final da década 1980 (Okerson, 1991), às iniciativas isoladas na década de 1990 tal como o arXiv (1991) de Paul Ginsparg e a Proposta Subversiva de Steven Harnad (1994) lançada em grupos de discussão e eventos científicos. Mas, foi somente na década de 2000 que houve um alinhamento de diretrizes e ações coordenadas que o acesso aberto ganhou força e destaque na comunidade internacional sensibilizando diversos países dos hemisférios de Norte a Sul.

Estas ações ganharam força e notoriedade à medida que pesquisadores de todo o mundo se engajaram em favor do AA. A assinatura de declarações e manifestos em apoio ao movimento foi uma das ações mais praticadas seguida de diretrizes institucionais, nacionais e internacionais para promover as estratégias do acesso aberto em nível mundial.

Tão logo surgem as iniciativas em Acesso Aberto - AA, as boas novas são acolhidas pelo Brasil em forma de ações estratégicas para o país, especialmente pelo Instituto Brasileiro de Ciência e Tecnologia - IBICT que se inseriu imediatamente ao Movimento do Acesso Aberto. Dentre essas ações, o IBICT desenvolveu parcerias, inclusive com agências de fomento, que viabilizaram recursos para projetos voltados para a implantação de repositórios, bem como esforços em larga escala para transformar os periódicos impressos em eletrônicos, visando a consolidação do AA no Brasil.

Apesar das diversas iniciativas nacionais para o desenvolvimento das duas estratégias propostas pela BOAI (AA Verde e AA Dourado, hoje denominado de Diamante), a consolidação do AA vem ocorrendo de formas distintas, especialmente em países em desenvolvimento, particularmente na América Latina e Caribe conforme a literatura da área vem evidenciando.

As características de cada área do conhecimento é um dos fatores de desigualdade mais expressivos. Há estudos clássicos, como o de Line (1981), que comprovaram que cada área apresenta seu perfil próprio, tal como as ciências sociais entre outros estudos mais contemporâneos publicados no país (Furnival; Guirra, 2017; Furnival; Silva-Jerez, 2017; Costa; Weitzel; Leta, 2020). Este fator, denominado por Ortellado (2008) de "desigual padrão", gera impactos distintos no sistema de produção científica, afetando, portanto, o "nível" de plenitude do Acesso Aberto nessas áreas. Essa desigualdade pode, por exemplo, reforçar as estratégias de pagamento de taxas para publicar artigos em Acesso Aberto nas áreas das Ciências da Saúde exigidos por editores comerciais que concentram revistas de maior fator de impacto; diferentemente da Letras, Linguística e Artes que é uma área que apresentou total adesão ao AA Diamante no Brasil pois muitos títulos são publicados por instituições acadêmicas reconhecidas no país caracterizando-se como editoras não-comerciais.

Dessa forma, além das diferenças naturais das áreas do conhecimento, há também as formas diferenciadas no desenvolvimento e consolidação do AA em países da América Latina e Caribe, tal como destacaram Leite e Costa (2016). Para os autores, esta diferença está na base das características do sistema de comunicação científica da região, estabelecida ao longo do tempo. Um bom exemplo disso foi citado por Castro (2015) que considera que a forte presença do Estado no financiamento de pesquisas científicas por meio das agências de fomento, influenciou intensamente na qualidade e na quantidade da produção científica nacional.

Mas, apesar da forte tendência do AA Diamante na América Latina e Caribe ser referência, devido à natureza não-comercial da maioria dos editores que publicam na região, o mesmo não foi suficiente para impulsionar o AA Verde (Vessuril; Guédon; Cetto, 2014), especialmente em áreas que publicam em periódicos comerciais com alto fator de impacto.

Sobre essa situação Ortellado (2008) acredita que “o país descuidou das políticas” governamentais para o desenvolvimento e maior adesão do AA Verde, apesar dos esforços inovadores do IBICT, justamente devido ao reconhecimento do sucesso do AA Dourado, atualmente denominado de AA Diamante, no cenário brasileiro e latino-americano e caribenho por iniciativas de destaque como o SciELO. Assim, embora o país lidere o ranking mundial de artigos periódicos científicos em Acesso Aberto, onde a publicação “nasce” disponível, sem barreiras de pagamento por ambas as partes, leitor e autor, existe ainda contextos em que a publicação está sob o acesso restrito sendo, portanto, fundamental a estratégia do AA Verde para torná-la disponível amplamente.

Por outro lado, de um modo geral, a produção científica brasileira difere da produção da ciência mundial por possuir características específicas e singulares que envolvem o modelo de Comunicação Científica brasileira. Para Spagnolo (1989) a pesquisa nacional está inteiramente relacionada às universidades e pós-graduações brasileiras. Portanto, é comum o predomínio

no Brasil de periódicos e editoras vinculados a estas instituições. Outra característica apontada por Castro (2015) que indica uma forte presença do estado no financiamento de pesquisas científicas através dos órgãos de fomento, acarretando uma forte influência na produção científica nacional que muitas vezes é escoada por publicações estrangeiras. Para Furnival (2015), os pesquisadores brasileiros têm preferência por publicarem em periódicos com alto fator de impacto, pois além de terem mais prestígio, são condições para obter financiamentos e verbas de órgãos públicos de fomento, que dão enfoque em suas avaliações a dados quantitativos. Tais anacronismos podem afetar o desenvolvimento do Acesso Aberto no país.

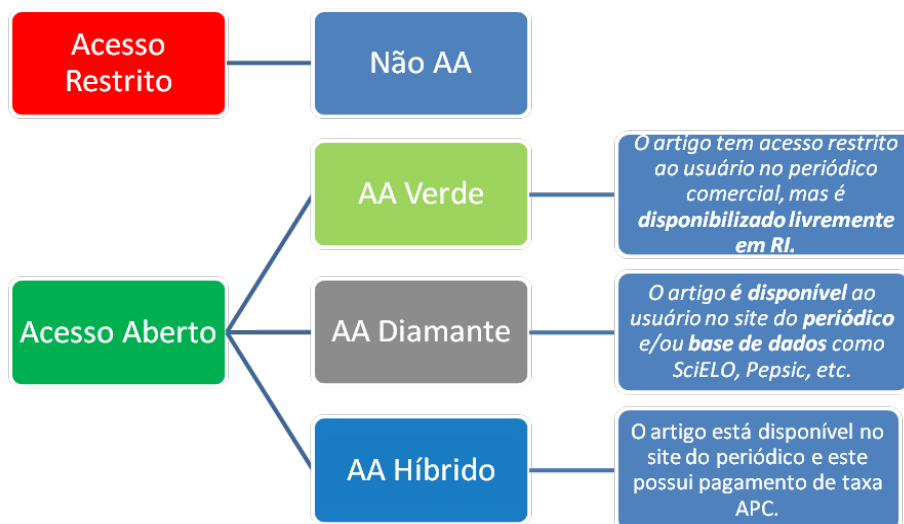
Por isso, o presente estudo está focado no AA Verde brasileiro e tem por objetivo verificar as suas principais características por meio de três estudos de casos representativos a fim de levantar aspectos de relevância que possam fomentar e orientar as estratégias em larga escala para o país.

Este estudo faz parte de uma análise mais ampla, ainda em desenvolvimento, da observação de cada área do conhecimento, conforme a Tabela de Áreas do Conhecimento/Avaliação da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES e sua relação com Acesso Aberto para assim caracterizar o Acesso Aberto Verde no Brasil.

Nesta etapa do estudo, foram examinadas as 3 áreas, a saber: Ciências da Saúde, Letras/Linguística e Engenharias, representados respectivamente pelos cursos de Pós-graduação em Medicina/Biomedicina, Letras e Engenharia da UFRGS, UnB e UFBA. A seleção dos casos foi baseada no The World University Rankings 2022 e sorteadas aleatoriamente a partir das 20 melhores posições e que possuem os três programas de pós-graduação, no ano de 2021.

PROGRAMA/GRANDE ÁREA /ÁREA	CIÊNCIAS DA SAÚDE	ENGENHARIAS	LETRAS, LINGÜÍSTICAS E ARTES
UnB	Medicina Tropical	Engenharia Elétrica	Linguística
UFBA	Medicina	Engenharia	Letras
UFRGS	Medicina	Engenharia Civil	Letras

Os artigos de periódicos das IES pesquisadas foram selecionados através de pesquisa na Plataforma Sucupira, por meio dos dados do Coleta Capes do ano de 2021 e 2022, sorteados aleatoriamente em site sorteador e agrupados em Acesso Aberto e Acesso Restrito. Destas duas categorias elaboradas, foram elencadas em 3 tipos de acesso a saber: AA Verde, AA Diamante e AA Híbrido.



Fonte: Autoras (2023)

A amostra da produção científica em levantamento realizado na Plataforma Sucupira, resultou nos seguintes dados: 582 artigos em Ciências da Saúde, 1.003 em Letras, Linguística e Artes e 168 nas Engenharias, dos quais foram analisados aproximadamente 5 % do total publicado. Os gráficos abaixo demonstram os dados por disciplina, caracterizando as áreas do conhecimento que são foco do presente estudo.

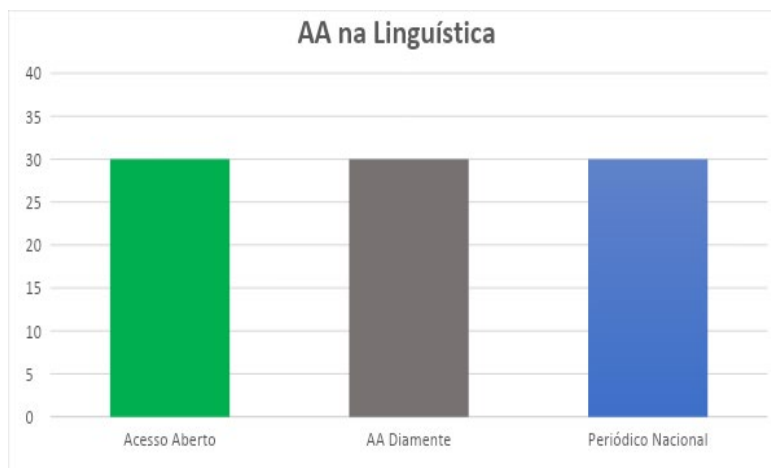


GRÁFICO 1. Apresenta os dados das Letras, Literatura e Artes.

Fonte: Autoras (2023)

O Gráfico 1 demonstra que os casos da área de Letras, Linguística e Artes, apresentou 100 % de sua produção em AA Diamante, com uma produção publicada exclusivamente em periódico nacional e no idioma nativo. Este resultado demonstra que a área vem investindo em periódicos de Acesso Aberto garantindo grande representatividade e relevância no AA Diamante, destacando-se também pelo maior número de publicações em periódicos nacionais. Também foi observado a baixa adesão a periódicos com pagamento de taxas e ausência de expressividade no Acesso Aberto Verde.

Nos casos selecionados da área das Engenharias, foi observado que há paridade em relação aos tipos de acesso onde 50 % da produção estão em Acesso Aberto e 50 % em Acesso Restrito. Foi observado também que a área



se apresenta bastante internacionalizada e que seus pesquisadores publicam, em sua maioria, em periódicos estrangeiros, totalizando 85 %. O AA Verde não foi praticado nos casos selecionados para representar as Engenharias, conforme apresentado no gráfico 2. Por outro lado, há forte tendência ao AA Híbrido, totalizando 53 % do total. O AA Diamante igualmente apresenta forte tendência na área ficando com 47 % do total.

Os gráficos abaixo apresentam os dados das Engenharias.

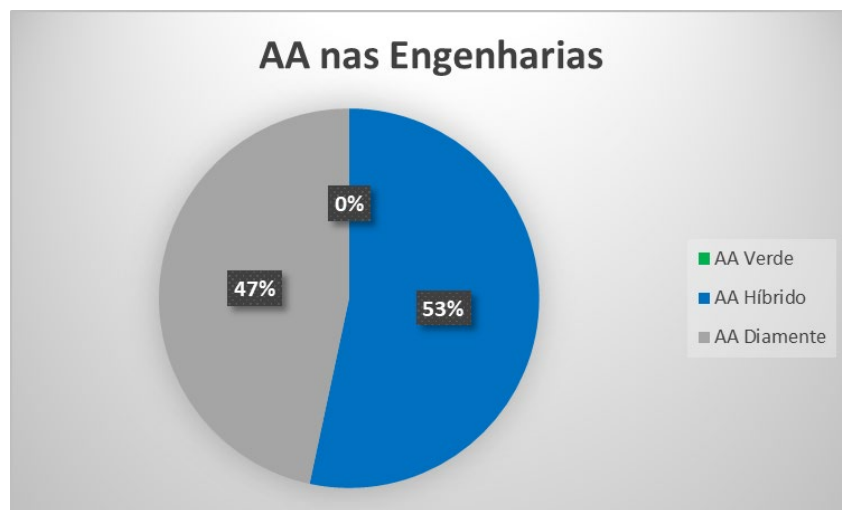


GRÁFICO 2. Acesso Aberto nas Engenharias

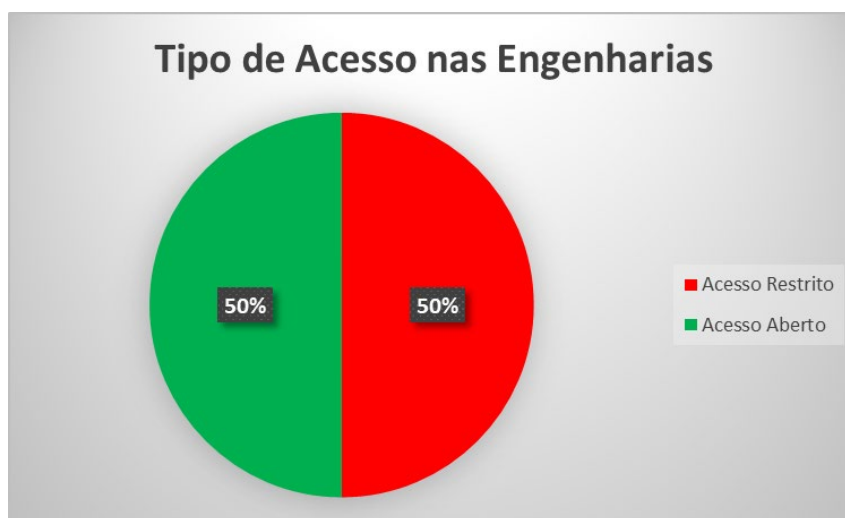


GRÁFICO 3. Tipo de Acesso nas Engenharias

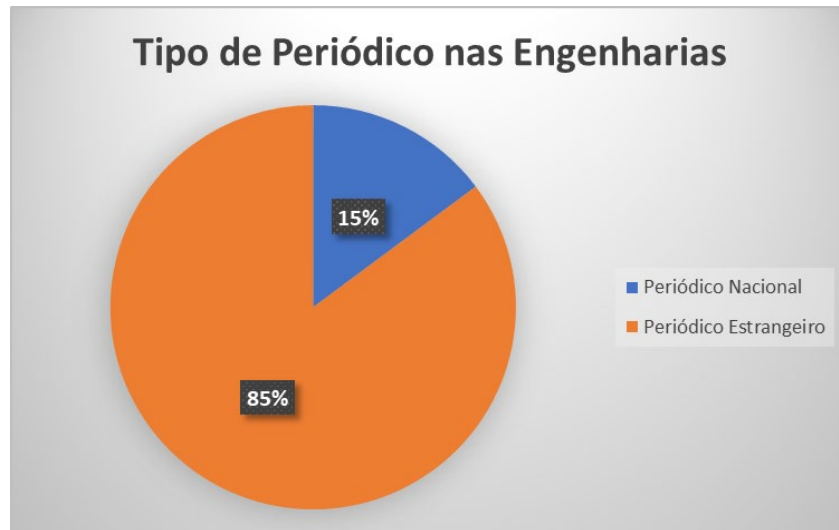


GRÁFICO 4. Tipo de Periódico nas Engenharias

Fonte: Autoras (2023)

Os casos que representam as Ciências da Saúde foram os únicos neste estudo, que demonstraram, mesmo que timidamente, um percentual de publicações disponíveis em AA Verde, aproximadamente 9% conforme pode ser observado no gráfico 5. O mesmo gráfico demonstra também uma adesão considerável dos pesquisadores ao AA Diamante com 48%, com desempenho igualmente constatado no AA Híbrido (43%). O fato do bom desempenho em AA Diamante, se deve ao fato da forte presença da área da saúde no SciELO e por periódicos nacionais, indexados em bases de dados estrangeiras.

Os gráficos abaixo apresentam os dados das Ciências da Saúde.

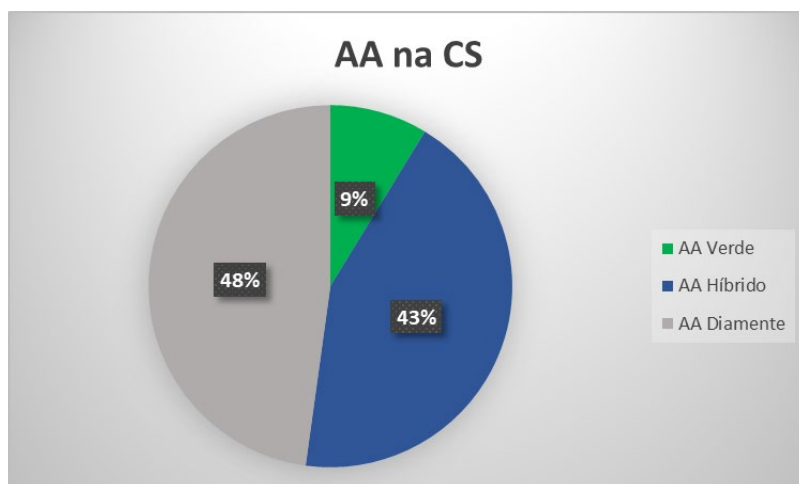


GRÁFICO 5. Acesso Aberto nas Ciências da Saúde

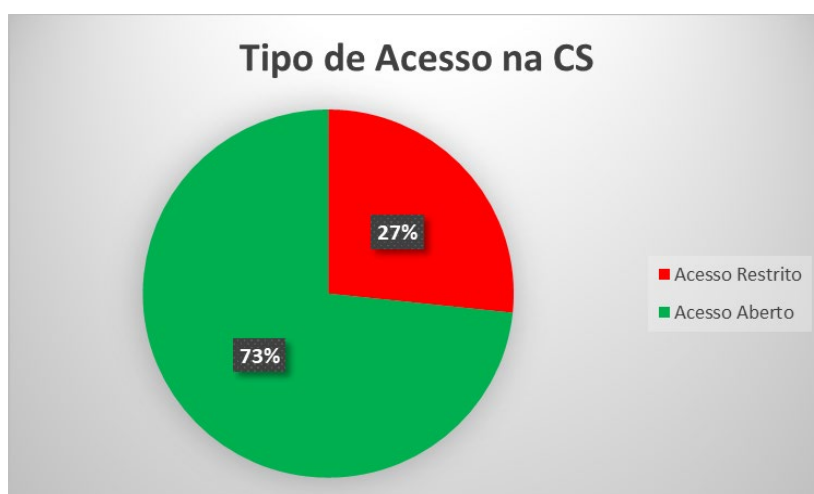


GRÁFICO 6. Tipo de Acesso nas Ciências da Saúde

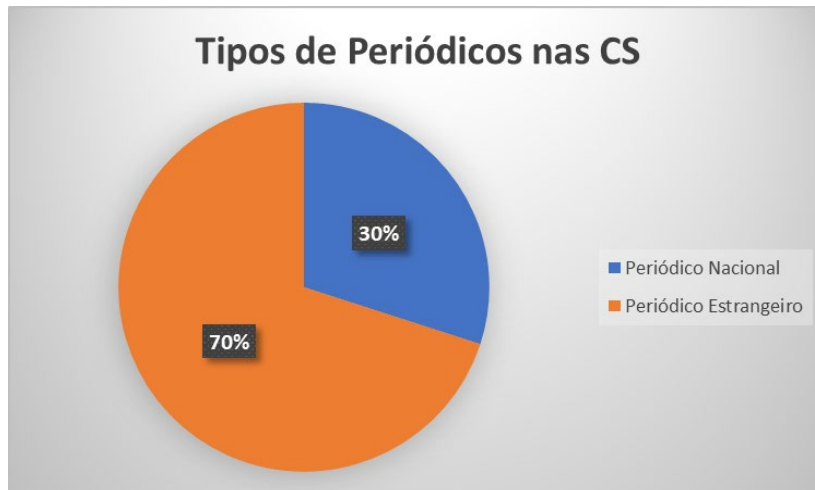


GRÁFICO 7. Tipo de periódico nas Ciências da Saúde

Fonte: Autoras (2023)

Este panorama, considerando as áreas do conhecimento apontadas neste estudo, conclui que cerca de 74 % da produção está disponível em Acesso Aberto e 26 % em Acesso Restrito, gráfico 10. Destes 75 % em AA, temos uma discreta adesão no AA Verde, menos de 3 %, e uma expressiva participação no AA Diamante, com um pouco mais de 70 %, gráfico 8. O AA Híbrido está representado com 26 %. O estudo demonstrou ainda uma preferência por periódicos estrangeiros, cerca de 51 %, gráfico 9, dados estes que acreditamos, foram alavancados pelas Engenharias.

Se levarmos em consideração que 25 % das publicações dos casos analisados, publicadas pelas IES brasileiras inseridas neste estudo estão em Acesso Restrito e que 26 % que estão em AA advém do AA Híbrido, podemos concluir que mais da metade das publicações, tem a necessidade de pagamento para se obter o acesso, seja por parte do leitor, seja pelo autor/instituição.

Os gráficos abaixo demonstram práticas de AA nãstrês áreas dos casos estudados.

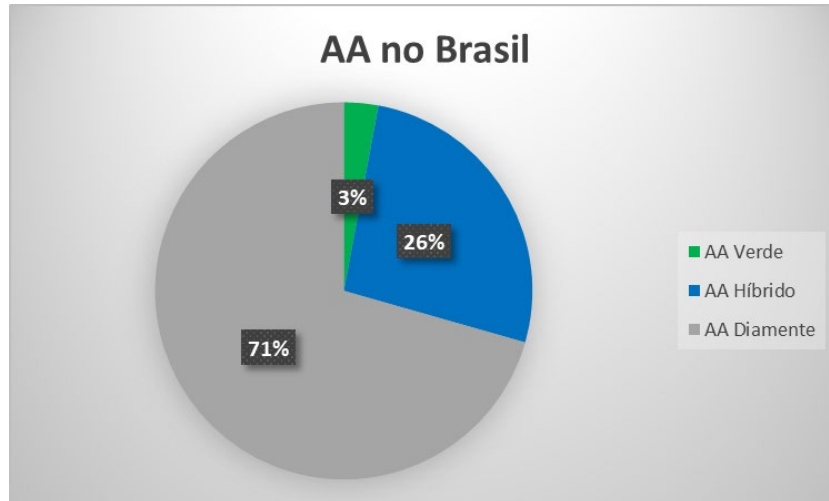


GRÁFICO 8. Acesso Aberto em geral

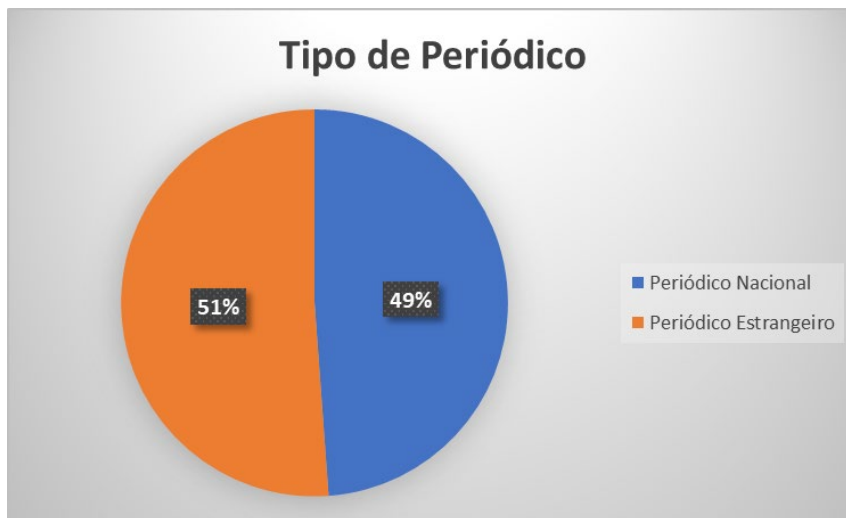


GRÁFICO 9. Tipo de Periódico em Geral

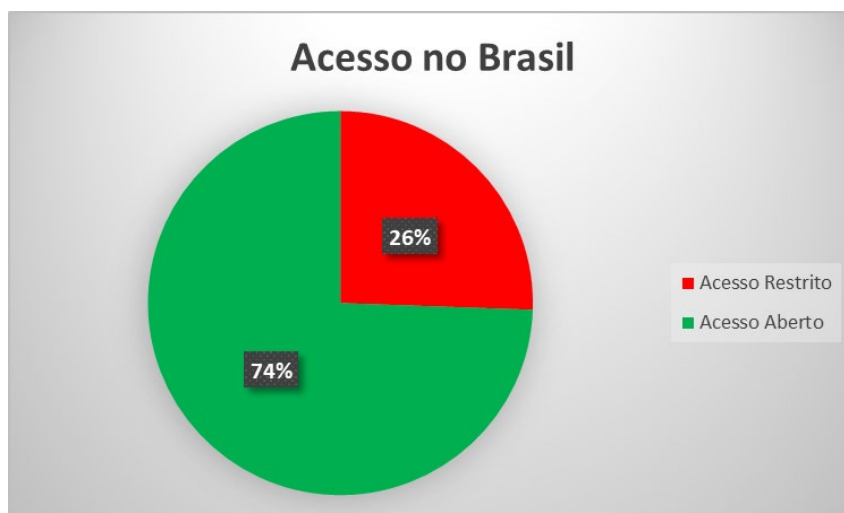


GRÁFICO 10. Tipo de acesso no geral

Fonte: Autoras (2023)

Apesar da importância da estratégia do AA Verde, especialmente para as áreas que publicam em periódicos comerciais de acesso restrito, os dados demonstraram que esta estratégia no Brasil não tem sido adotada de maneira plena nos termos defendidos pela BOAI (2002; 2012; 2022) e por Harnad (2004; 2012), uma vez que cerca de 25 %, gráfico 10, da produção dos casos selecionados nas áreas analisadas por este estudo, se encontra em Acesso Restrito, ou seja, os RI's brasileiros não estão cumprindo as recomendações do AA Verde que tem a premissa de disponibilizar os artigos publicados em periódicos de assinatura.

### Considerações finais

O monitoramento dos artigos de periódicos dos casos selecionados por tipo de acesso apresentados neste estudo, relacionados com os demais estudos na área, podem refletir as políticas nacionais de fomento à pesquisa brasileira que incentivam a internacionalização dos periódicos científicos (Vogel; Kobashi, 2015) bem como todo o esforço dos pesquisadores em qualificar os periódicos nacionais. Por outro lado, para autores como Guédon (2010) o predomínio do

AA Diamante no Brasil pode indicar características próprias de ciência periférica.

De acordo com Furnival (2015) os pesquisadores no Brasil optam por publicar em periódicos de alto fator de impacto em função das cobranças institucionais e, sobretudo, das avaliações de desempenho tal como foi comprovado no estudo com a elite de pesquisadores no país (Costa; Weitzel; Leta, 2020). Furnival (2015) também destaca a vocação de pesquisa de ponta do país em certas áreas do conhecimento que podem ser consideradas como mainstream, apesar de suas características de país de periferia científica.

Apesar desses anacronismos, há atualmente no Brasil, assim como na América Latina e Caribe, um ambiente favorável por parte das Instituições de Ensino, Pesquisa e outras instituições de fomento, para a adoção de práticas que favoreçam um ecossistema em AA, com implementação inclusive de plataformas comunitárias, tal como La Referência, para promoção o AA.

O maior desafio nacional, que este estudo pode observar e que também está descrito como prioridade na BOAI (2022), são: as práticas de avaliação de desempenho do pesquisador que potencializam a desigualdade entre as áreas do conhecimento; o grande volume de publicação em AA Híbrido, por motivo igualmente destoantes entre as áreas; e por fim, a baixa adoção do AA Verde no cenário nacional, onde mesmo em um cenário relevante de implementação de RI's, estes, como este estudo observou, possuem em sua maioria, publicações que já "nasceram" em Acesso Aberto.

Este estudo sugere ainda medidas como: a) mudanças na prática de avaliação, como os Fatores de Impacto ou o Sistema Qualis (Sistema nacional de avaliação de periódicos). Há alternativas viáveis mais justas como o exemplo da Universidade de Liege (Bélgica) que atrela a progressão de carreira às publicações que seus pesquisadores têm no RI, além do currículo que conste suas publicações (Santos; Kishi, 2015); b) buscar o distanciamento do AA Híbrido, não com a mesma demanda de adoção de medidas, como vem

ocorrendo em países desenvolvidos a exemplo do Plano de Ação Diamond Open Access, uma vez que o cenário nacional já contempla a prevalência do AA Diamante. É, portanto, de vital importância retomar a “Evangelição”, termo adotado por Harnad, entre os pesquisadores para adesão ao Movimento do Acesso aberto, com conscientização das práticas que envolvem o mercado comercial científico e suas práticas predatórias; c) buscar o fortalecimento do AA Verde com depósito em RI nacionais, uma vez que a pesquisa preliminar identificou, que o depósito ficou aquém do grande papel que o Brasil vem desempenhando no AA Diamante de forma a contribuir mais plenamente para o acesso aberto amplo e irrestrito.

A adesão do país, através da Capes, aos chamados “Acordos Transformativos” (AT’s), é outro aspecto bastante preocupante, pois pode erradicar o Acesso Aberto Verde no país<sup>1</sup>. Desde 2000 a Capes mantém um Portal de Periódicos, que disponibiliza acesso para as comunidades acadêmicas com financiamento governamental de conteúdos de acesso restrito. De 2000 a 2020 o valor acumulado pago pelo governo brasileiro às editoras estrangeiras totalizou aproximadamente, impressionante \$ 1.231.755.000,00. Segundo Alencar e Barbosa (2022) foram gastos com APC’s no Brasil, no período de 2012 a 2019, em média, U\$ 631,39 por artigo de periódico. Esse valor passou para U\$ 1.131,35 em 2019, apresentando um aumento de 79%. Nesse mesmo período, autores e financiadores de pesquisa brasileiros gastaram mais de quarenta milhões de dólares em APC’s. O estudo ainda indica uma tendência real de publicar em periódicos mais caros e em quantidade significativa.

---

<sup>1</sup> Para a BOAI (2022), esse modelo, denominado de “ReadandPublish” ou “PublishandRead” ou ainda “acordos de compensação”, não é transformador, pelo contrário, mantém e valoriza os monopólios das grandes editoras comerciais e suas ações predatórias. Tais acordos envolvem contratos e licenças de assinaturas de periódicos de editoras comerciais que garantem alguns descontos em taxas para publicar em suas revistas que as instituições de pesquisa e acadêmica assinam ou fundações como a Capes no Brasil.



Sem acesso ao Portal de Periódicos da Capes haverá disparidade no acesso igualitário e irrestrito à uma parcela da população, além dos altos custos de financiamento para pagamento de assinaturas de periódicos comerciais. Enquanto que nos demais países do mundo, onde os periódicos comerciais tem predomínio, estão se buscando alternativas, mesmo que duvidosas, de acesso aberto total (full open access), a Capes parece ignorar a tradição nacional do mercado (Ramalho; Silva; Rocha, 2020).

Apesar de ser animador o fato de instituições estarem se mobilizando para encontrarem alternativas viáveis para o AA total, de acordo com a edição 2022 da BOAI:

... devemos favorecer modelos de publicação que beneficiem todas as regiões do mundo, que sejam controlados por organizações acadêmicas e sem fins lucrativos, que evitem concentrar a produção científica de acesso aberto em periódicos comercialmente dominantes e que evitem modelos entrincheirados que conflitem com esses objetivos. Recomendamos não fazer acordos de "compensação", "ler e publicar" ou os chamados acordos "transformativos (BOAI, 2022).

Portanto, ainda há um longo caminho para analisar os impactos da adoção de AT's pela comunidade científica, especialmente em relação aos esforços que pesquisadores e instituições investiram para alavancar o Acesso Aberto no país. Mas, de uma forma ou de outra, o país ainda carece de recursos financeiros e estratégias mais inovadoras para transformar o sistema de publicação científica como um todo bem como suas práticas científicas visando o acesso mais justo e igualitário.

## **Bibliografía**

ALENCAR, B. N.; BARBOSA, M. C. (2021). Publicações de acessoaberto com pagamento por Article Processing Charge (APC): uma análise do

- cenário brasileiro. *Anais da Academia Brasileira de Ciências*, 93(4),1984. DOI: <<https://doi.org/10.1590/0001-3765202120201984>>
- BRUMATTI, J. D. (2016). O Acesso Aberto Verde no Brasil: umestudodescritivo da produção científica depositada em Repositório Institucional. [Dissertação de Mestrado]. Universidade Federal do Estado do Rio de Janeiro.
- BUDAPESTE OPEN ACCESS INITIATIVE. (2002). Iniciativa de Budapeste pelo AcessoAberto. Budapeste. <<http://www.budapestopenaccessinitiative.org/translations/portuguese-translation>>
- BUDAPESTE OPEN ACCESS INITIATIVE. (2012). Dez anos da iniciativa de Budapeste em AcessoAberto: a abertura como caminho a seguir. Budapeste. <<http://www.budapestopenaccessinitiative.org/boai-10-translations/portuguese-braziliantranslation>>
- BUDAPESTE OPEN ACCESS INITIATIVE. (2022). TheBudapeste Open Access Initiative: 20th anniversaryrecommendations. Budapeste. <<https://www.budapestopenaccessinitiative.org/boai20/>>
- CASTRO, A. L da S. (2015). Ascensão da produção científica brasileira: risco iminente de um colapso? *Multi-ScienceJournal*, 1(2), 1-2.
- COSTA, E. H. S.; WEITZEL, S. DA R.; LETA, J. (2020). Adesão da elite brasileira de pesquisadoresaos periódicos de acessoaberto: a relaçãocomgênero, região geográfica e grande área do conhecimento. *Questão*, 26(3), 15-42. DOI: <<https://doi.org/10.19132/1808-5245263.15-42>>
- FURNIVAL, A. C. (2015). Brasil é referência em acessoaberto, mas faltam políticas integradas [Entrevista concedida à Patrícia Santos e Kátia Kishi]. <<http://www.oei.es/historico/divulgacioncientifica/?Brasil-e-referencia-em-acesso>>
- FURNIVAL, A. C.; GUIRRA, D. A. R. (2017) As percepções e práticas de publicação em acessoaberto dos pesquisadores de dois programas de pós-graduação em engenharia. *Revista Digital de Biblioteconomia e Ciência da Informação*, 15(2), 469-488.
- FURNIVAL, A. C. M.; SILVA JEREZ, N. S. (2017). Percepções de pesquisadores brasileiros sobre o acessoaberto à literatura científica. *Informação & Sociedade: Estudos, João Pessoa*, 27(2), 153-166.
- GUÉDON, J-C. (2010). AcessoAberto e divisão entre ciência predominante e ciência periférica. In: *Acessibilidade e visibilidade de revistas científicas eletrônicas*. Ferreira, S. M. S. Pinto; Targino, M. G. (Org.). São Paulo: Senac São Paulo; CENGAGE Learning, p. 21-73.
- HARNAD, S. (1994). Publiclyretrievable ftp archives foresotericscience and scholarship: a subversive proposal. <<https://wiki.lib.sun.ac.za/images/e/ef/A-subversive-proposal.pdf>>
- HARNAD, S., BRODY, T., VALLIERES, F., CARR, L., HITCHCOCK, S., GINGRAS, Y., OPPENHEIM, C., STAMERJOHANN, H., & HILF, E. (2004). Thegreen and the

- gold roads to Open Access. Nature Web Focus.  
<<https://eprints.soton.ac.uk/259940/>>
- HARNAD, S. (2012a). Hybridgold open access and thechesirecat'sgrin: howtorepairthe new open accesspolicyof RCUK. LSE Impactof Social Sciences. <<http://eprints.soton.ac.uk/342582/>>
- HARNAD, S. (2012b). Public Access toFederallyFundedResearch (Harnad Response to US OSTP RFI) (Part 1 of 2). Open Access Archivangelism. <<https://web-archive.southampton.ac.uk/openaccess.eprints.org/index.php%3F%252Farchives%252F865-guid.html>>
- HARNAD, S. (2012c). Whythe UK shouldnotheadtheFinchReport. Impactof Social Sciences blog. <<https://blogs.lse.ac.uk/impactofsocialsciences/2012/07/04/why-the-uk-should-not-heed-the-finch-report/>>
- LEITE, F. C. L., & COSTA, S. M. S. (2016). Modelo genérico de gestão da informação científica para instituições de pesquisa na perspectiva da comunicação científica e do acessoaberto. *Investigación Bibliotecológica: Archivonomía, Bibliotecología e Información*, 30(69), 43-73.
- LINE, M. B. (1981). Thestructureof social scienceliterature as shownby a large-scalecitationanalysis, *Social Science Information Studies*, 1(2), 67-87. DOI: <[https://doi.org/10.1016/0143-6236\(81\)90001-6](https://doi.org/10.1016/0143-6236(81)90001-6)>
- OKERSON, A. (1991). Theelectronicjournal: what, whence, and when? *The Public-Access Computer Systems Review*, 2(1), 5-24. <[https://uh-ir.tdl.org/bitstream/handle/10657/5146/Okerson\\_1991\\_TheElectronicJournal.pdf?sequence=1&isAllowed=y](https://uh-ir.tdl.org/bitstream/handle/10657/5146/Okerson_1991_TheElectronicJournal.pdf?sequence=1&isAllowed=y)>
- ORTELLADO, P. (2008). As políticas nacionais de acesso à informação científica. *LIINC em Revista*, 4(2), 186-195.
- PLATAFORMA SUCUPIRA. (2023). Coleta Capes. <<https://sucupira.capes.gov.br/sucupira/>>
- SANTOS, P.; KISHI, K. (2015). Brasil é referência em acessoaberto, mas faltam políticas integradas. ComCiência. <<https://comciencia.br/comciencia/handler.php?section=8&edicao=11&id=1339>>
- THE TIMES HIGHER EDUCATION. (2022). World University Rankings. <<https://www.timeshighereducation.com/contact-us>>
- UNIVERSIDADE FEDERAL DA BAHIA. (2023). Repositório Institucional da UFBA. <<https://repositorio.ufba.br/>>
- UNIVERSIDADE FEDERAL DE BRASÍLIA. Repositório Institucional da UnB. <<https://repositorio.unb.br/>>
- UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL. LUME Repositório Digital. <<https://www.lume.ufrgs.br/>>

- VESSURI, H., GUÉDON, J.-C., & CETTO, A. M. (2014). Excellence or quality? Impact of the current competition regime on science and scientific publishing in Latin America and its implications for development. *Current Sociology*, 62(5), 647-665. DOI: <https://doi.org/10.1177/0011392113512839>
- VOGEL, M., & KOBASHI, N. (2015). Avaliação da pós-graduação no Brasil: seus critérios. In *Anais do 16º Encontro Nacional de Pesquisa em Ciência da Informação* (p. 1-8). João Pessoa, PB.

## e-compartilhando

### **DANIELA DANIGNO KONRADT**

IFSUL

[danielakonradt@ifsul.edu.br](mailto:danielakonradt@ifsul.edu.br)

### **VANESSA MATTOSO CARDOSO**

IFSUL

[vanessacardoso@ifsul.edu.br](mailto:vanessacardoso@ifsul.edu.br)

### **TIAGO DA SILVA GAUTERIO**

IFSUL

[tiagogauterio@ifsul.edu.br](mailto:tiagogauterio@ifsul.edu.br)

### **ADRIANE RODRIGUES CORREA**

IFSUL

[adrianecorrea@ifsul.edu.br](mailto:adrianecorrea@ifsul.edu.br)

### **WALKIRIA CORDENONZI**

IFSUL

[walkiriacordenonzi@ifsul.edu.br](mailto:walkiriacordenonzi@ifsul.edu.br)

### **EJE TEMÁTICO**

Datos abiertos

### **RESUMEN**

A presente proposta tem como objetivo estabelecer uma comunidade de prática e um repositório colaborativo entre o IFSul Câmpus SL/ Brasil, e instituições parceiras Uruguai, com o intuito de fortalecer a formação de pesquisadores, promover a produção de materiais científicos e realizar workshops colaborativos. Essas parcerias são fundamentais para garantir a participação ativa dos pesquisadores, bem como o compartilhamento de

recursos e conhecimentos entre as instituições. A metodologia proposta envolve a criação de uma plataforma online, servindo como base para a comunidade de prática e para o repositório colaborativo. Essa plataforma está sendo desenvolvida com recursos de acesso aberto e tecnologias de compartilhamento de documentos, permitindo a interação e colaboração entre os membros da comunidade. Para a formação de pesquisadores, estão sendo organizados workshops presenciais e virtuais, com a participação de especialistas tanto do IFSul quanto das instituições parceiras, abordando tópicos relevantes para a pesquisa nas áreas de interesse comum, proporcionando a troca de conhecimentos e experiências entre os participantes. Quanto à produção de materiais científicos, os pesquisadores serão incentivados a compartilhar seus resultados e trabalhos acadêmicos no repositório colaborativo, inclusive com os artigos que já foram publicados pelos mesmos. Isso garantirá o acesso aberto a esses materiais e, conseqüentemente, ao conhecimento. Em suma, a criação dessa comunidade de prática e desse repositório colaborativo proporcionará uma plataforma dinâmica para a formação de pesquisadores, a produção de materiais científicos e a realização de workshops colaborativos. Essa iniciativa impulsionará a colaboração internacional, o intercâmbio de conhecimentos e contribuirá para o avanço científico nas áreas de interesse mútuo.

#### **PALABRAS CLAVES**

Comunidade de prática; repositório colaborativo; formação de pesquisadores; dados abertos.

# Práticas de compartilhamento de dados pelos pesquisadores PQ da área de Física e Astronomia do CNPQ

**RENE FAUSTINO GABRIEL JUNIOR**

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

[rene.gabriel@ufrgs.br](mailto:rene.gabriel@ufrgs.br)

## EJE TEMÁTICO

Datos abiertos

## RESUMO

Compartilhar os dados de pesquisa significa promover a transparência e possibilitar a reprodutibilidade das pesquisas. Faz-se a importância de estudar as práticas de compartilhamento de dados de pesquisa pelas diversas áreas do conhecimento para compreender suas especificidades e práticas. Este estudo tem como objetivo identificar as práticas de compartilhamento de dados por pesquisadores da área de física e astronomia da CAPES. Para tal, utiliza-se ferramentas de scraping para coleta dos pesquisadores PQ, totalizando o número de 1.078 autores. Para possibilitar a análise foram coletadas das produções científicas, foi utilizando o XML do LattesExtrator do CNPq, resultando em 107.444 artigos, sendo 19.377 (18,0 %) sem DOI. Do restante, 88.607 (82,0 %) foi consultado via API em Python o ScholeXplorer para verificar a existências de relacionamentos com outras publicações ou datasets. Resultando em 4.256 relacionamento de 3.483 (16,7 %) das publicações. O ScholeXplorer é uma plataforma aberta que permite estabelecer uma estrutura de interoperabilidade de alto nível para troca de informações sobre os links entre literatura acadêmica e dados. Observou-se que a prática da área da física está na disponibilização de dados em repositórios preprint como arXiv, mas também foram localizados depósitos de dados no CCDC, Figshare, Zenodo, Mendlay Data, entre outros em menor frequência. Considera-se que a foi possível compreender a área pelo cruzamento de DOIs de publicação e dados, porém precisaria testar com outras áreas do conhecimento.

**PALABRAS CLAVES**

Repositórios de dados de pesquisa; dados de pesquisa; ScholeXplorer.



# Geração de Valor a partir da Oferta de Repositórios Institucionais de Dados de Pesquisa

## **CAROLINA FELICISSIMO**

REDE NACIONAL DE PESQUISA

[carolina.felicissimo@rnp.br](mailto:carolina.felicissimo@rnp.br)

## **LEANDRO CIUFFO**

REDE NACIONAL DE PESQUISA

[leandro.ciuffo@rnp.br](mailto:leandro.ciuffo@rnp.br)

## **RAFAEL PORT DA ROCHA**

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

[rafael.rocha@ufrgs.br](mailto:rafael.rocha@ufrgs.br)

## **SÔNIA ELISA CAREGNATO**

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

[sonia.caregnato@ufrgs.br](mailto:sonia.caregnato@ufrgs.br)

## **RENE FAUSTINO GABRIEL JUNIOR**

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

[rene.gabriel@ufrgs.br](mailto:rene.gabriel@ufrgs.br)

## **CATERINA MARTA GROPOSO PAVÃO**

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

[caterina@ufrgs.br](mailto:caterina@ufrgs.br)

## **SAMILE ANDRÉA DE SOUZA VANZ**

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

[samile.vanz@ufrgs.br](mailto:samile.vanz@ufrgs.br)

## **MARIA LUIZA MACHADO CAMPOS**

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

[mluiza@ppgi.ufrj.br](mailto:mluiza@ppgi.ufrj.br)

## **WASHINGTON LUÍS RIBEIRO DE CARVALHO SEGUNDO**

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA

[washingtonsegundo@ibict.br](mailto:washingtonsegundo@ibict.br)

### **EIXO TEMÁTICO**

Datos abiertos

### **RESUMO**

Esse artigo apresenta como a oferta de repositórios institucionais de dados abertos de pesquisa pode gerar valor para produtores e consumidores de dados, além de gestores e patrocinadores do serviço. A experiência é retratada a partir do cumprimento de um Edital Público, lançado no segundo semestre do ano de 2021, fruto de um acordo de cooperação técnica celebrado entre a Rede Nacional de Ensino e Pesquisa (RNP), o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) e o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). As universidades e institutos de ciência e tecnologia (ICTs) selecionadas pelo edital realizaram atividades para a oferta de seus respectivos repositórios institucionais e relataram geração de valor durante todo o processo. O arcabouço para gerenciamento de serviços de tecnologia de informação baseado no ITIL@ 4 foi utilizado para demonstração dos resultados alcançados. A execução das atividades de pesquisa e desenvolvimento foram lideradas pelo Grupo de Trabalho Rede de Dados de Pesquisa Brasil (GT-RDP) da Universidade Federal do Rio Grande do Sul (UFRGS). Após o término do período previsto do edital, foi constituída uma rede colaborativa para acesso aberto de dados de pesquisa constituída de forma voluntária e participativa de seus membros.

### **PALAVRAS CHAVE**

Dados Abertos de Pesquisa; Repositórios Institucionais de Dados de Pesquisa; Rede de Repositórios Digitais. Open Research Data; Institutional Research Data Repositories; Digital Repositories Network.

## Introdução

Ao longo dos últimos anos, o Brasil tem participado em ações para promoção da Ciência Aberta. Em 2018, a Controladoria-Geral da União lançou um processo de elaboração do 4º Plano de Ação Nacional da Parceria para Governo Aberto com participação da sociedade civil na definição de temas por meio de consulta pública (Axelos, s. f.). O resultado do 4º Plano de Ação Nacional foi compilado com a proposição de onze compromissos, cocriados com o envolvimento de cento e cinco pessoas, representantes de oitenta e oito instituições, sendo trinta e nove organizações da sociedade civil, trinta e nove órgãos da Administração Pública Federal e dez órgãos das Administrações Públicas Estaduais e Municipais (Controladoria-Geral da União, 2014).

Dentre os onze compromissos propostos, o “Compromisso 3 – Inovação e Governo Aberto na Ciência” (Controladoria-Geral da União, 2018a) trouxe como objetivo: “Estabelecer mecanismos de governança de dados científicos para o avanço da ciência aberta no Brasil”. Tal compromisso foi construído, seguindo a metodologia da CGU, em duas oficinas de cocriação. As oficinas contaram com a participação paritária de órgãos públicos e da sociedade civil, cujos representantes são indicados pelo coordenador do compromisso, pelo Grupo de Trabalho da Sociedade Civil e pelo Grupo-Executivo do Comitê Interministerial para Governo Aberto (GE-CIGA) que é presidido pelo Ministério da Transparência e Controladoria-Geral da União (CGU).

Em Bertin e colegas (Controladoria-Geral da União, 2018b), o processo de construção do “Compromisso 3 – Inovação e Governo Aberto na Ciência” é apresentado além de como foram derivados seus respectivos marcos, mensuráveis e verificáveis para monitoramento e execução do compromisso, com os prazos de início e fim; e os órgãos responsáveis pela execução das ações de cada marco. Dentre os nove marcos propostos, os autores desse artigo foram responsáveis e/ou atuaram como protagonistas nos “Marco 7 - Implantação de infraestrutura federada piloto de repositórios de dados de pesquisa” e “Marco 8 - Proposição de padrões de interoperabilidade para

repositórios de dados de pesquisa”. Tal atuação foi impulsionada pela coordenação executiva da Rede Nacional de Ensino e Pesquisa (RNP) em parceria com o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), tendo a equipe acadêmica do Grupo de Trabalho Rede de Dados de Pesquisa Brasil (GT-RDP) (Bertin et al., 2019) da Universidade Federal do Rio Grande do Sul (UFRGS) como executores e validadores das ações realizadas.

Este artigo tem por objetivo relatar os avanços na área de Ciência Aberta alcançados no Brasil por meio da criação de uma rede embrionária de repositórios institucionais de dados de pesquisa através do projeto REDEDADOSABERTOS, executado pelo GT-RDP. REDEDADOSABERTOS é fruto de um Edital Público, lançado no segundo semestre do ano de 2021, através de um acordo de cooperação técnica celebrado entre a Rede Nacional de Ensino e Pesquisa (RNP), o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) e o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Tal edital visa a promoção da Ciência Aberta no Brasil por meio da sustentação do Consórcio Nacional para Ciência Aberta (CoNCienciA), liderado pelo CNPq.

### **Acesso Aberto a Dados de Pesquisa no Brasil**

Desde 2017, através da formalização de um Acordo de Cooperação Técnica (Universidade Federal do Rio Grande do Sul, 2022), RNP e IBICT vêm atuando, em conjunto, na promoção de atividades de pesquisa e desenvolvimento (P&D) em projetos nacionais na área de Acesso Aberto à Dados de Pesquisa. Uma de suas primeiras ações foi o lançamento de uma chamada aberta com o objetivo de identificação de práticas, além de prototipação de serviço, de acesso aberto a dados de pesquisa (Brasil, 2017). O grupo de trabalho selecionado foi o “Rede de Dados de Pesquisa Brasileira (GT-RDP Brasil)” (RNP, 2017a) constituído, na época, por professores e bolsistas dos Centros de Documentação e Acervo Digital de Pesquisa (CEDAP) e de Processamento de Dados (CPD), ambos da

Universidade Federal do Rio Grande do Sul (UFRGS), além do Grupo de Pesquisa em Gerenciamento de Informações do Centro de Ciências Computacionais (C3) da Universidade Federal do Rio Grande (FURG). RNP fomentou com recursos financeiros para a contratação dos professores e bolsistas do GT-RDP Brasil enquanto o IBICT seguia com os direcionadores para o P&D a ser executado.

Ao longo de dois anos de projeto (2018 – 2020), o GT-RDP Brasil executou ações de P&D, tanto nas vertentes conceitual quanto tecnológica, através de provas de conceitos (PoCs) para abertura de dados institucionais com experimentação em Dataverse (RNP, 2017b) e King (s. f.), software aberto desenvolvido para servir como repositório de dados de pesquisa. Os resultados de P&D originaram diversas publicações importantes e orientadoras para a comunidade brasileira que, na época, começava a experimentar repositórios institucionais de dados de pesquisa. Algumas dessas publicações do GT-RDP Brasil são pontuadas a seguir:

- Acesso Aberto a Dados de Pesquisa no Brasil - Repositórios brasileiros de dados de pesquisa: <<http://hdl.handle.net/10183/185138>>
- Acesso Aberto a Dados de Pesquisa no Brasil - Práticas e Percepções dos Pesquisadores: <<http://hdl.handle.net/10183/185195>>
- Acesso Aberto a Dados de Pesquisa no Brasil - soluções tecnológicas para compartilhamento de dados no Brasil: <<http://hdl.handle.net/10183/185126>>
- Acesso aberto a Dados de Pesquisa no Brasil: grupos nacionais no portal do RDA: <<http://hdl.handle.net/20.500.11959/1257>>
- Acesso Aberto a Dados de Pesquisa no Brasil: política para o repositório de dados de pesquisa: <<http://hdl.handle.net/20.500.11959/1263>>
- Acesso Aberto a Dados de Pesquisa no Brasil: identificadores persistentes para dados de pesquisa: <<http://hdl.handle.net/20.500.11959/1258>>

- Acesso Aberto a Dados de Pesquisa no Brasil: Dataverse: licenças de uso: <<http://hdl.handle.net/20.500.11959/1260>>
- Acesso Aberto a Dados de Pesquisa no Brasil: Dataverse Documentação: Ativação do DOI: <<http://hdl.handle.net/20.500.11959/1259>>
- Acesso Aberto a Dados de Pesquisa no Brasil: planejamento para implantação de comunidade produtora de dados para o repositório Rede de Dados de Pesquisa: <<http://hdl.handle.net/20.500.11959/1204>>
- Pacote de tradução do Dataverse para Português BR: <<https://github.com/RNP-dadosabertos/dataverse-language-packs>>
- Script de instalação do Dataverse + Shibboleth no CentOS: <<https://github.com/RNP-dadosabertos/dataverse-centos>>

Após o período de estudo do estado da arte dos panoramas nacional e internacional para acesso aberto a dados de pesquisa e PoCs realizadas com a experimentação do software Dataverse para repositório de dados de pesquisa, o GT-RDP acumulou o conhecimento necessário para ações de capacitação e promoção da Ciência Aberta no Brasil. RNP, IBICT e CNPq, então, propiciariam o que viria a ser o projeto REDEDADOSABERTOS.

### **Edital Público RNP-IBICT-CNPq para promoção da Ciência Aberta**

Em 01 de julho de 2021, por um acordo técnico científico, RNP, IBICT e CNPq lançaram um edital público para promoção da Ciência Aberta Brasileira, no qual convidava Instituições de Ciência, Tecnologia e Inovação (ICTs) a colaborarem em um projeto de Pesquisa e Desenvolvimento (P&D) através de ações de transferência de conhecimento e incubação de repositórios de dados de pesquisa. ICTs são caracterizadas como órgãos ou entidades de administração pública ou privada, sem fins lucrativos, cuja principal missão institucional é desenvolver e executar atividades de pesquisa (básica ou aplicada) de caráter

científico ou tecnológico. Dentre exemplos de ICTs podemos citar as universidades federais e institutos nacionais de pesquisa.

As quatro ICTs selecionadas no edital para incubação de repositórios de dados de pesquisa foram: Universidade Federal Rural do Amazonas (UFRA), Universidade Federal de Goiás (UFG), Universidade Federal do Ceará (UFG) e o Centro Brasileiro de Pesquisas Físicas (CBPF) (King, 2007). As instituições receberam apoio na criação de seus repositórios de dados de pesquisa, por meio de ações de capacitação e transferência de conhecimento, durante um período de um ano. O aprendizado beneficiou essas instituições como um todo, em especial as bibliotecas de ensino e pesquisa, que foram o público-alvo da chamada.

Para entregar valor para as ICTs selecionadas (CBPF, UFRA, UFG e UFC), o GT-RDP Brasil avaliou o arcabouço para gerenciamento de serviços de tecnologia de informação baseado no ITIL@ 4. Dentre as propostas do ITIL@ 4 está a instanciação de uma Cadeia de Valor de Serviço (CVS), onde, para cada necessidade de entrada, são executadas atividades de forma a se chegar em valor como resultado.

### **A Cadeia de Valor de Serviço do ITIL@ 4**

ITIL@ 4 vem sendo utilizado mundialmente como um conjunto de práticas que podem ser aplicadas na direção da transformação digital para serviços de tecnologia de informação. Trata-se de um modelo não prescritível, independente de tecnologia e de domínio público, que permite a criação conjunta de valor entre produtores e consumidores. A partir de uma entrada, seja demanda ou oportunidade, uma série de atividades são executadas, gerando resultados que trazem benefícios percebidos por consumidores, produtores, patrocinadores, partes interessadas.

A Fig.1 ilustra a proposta da Cadeia de Valor de Serviço (CVS) apresentada pelo ITIL@ 4. Em linhas gerais, uma CVS é instanciada por uma Necessidade,

oriunda de uma oportunidade ou demanda de um cliente. A atividade inicial da CVS é denominada “Engajar”. A execução da atividade “Engajar” estimula a execução das outras atividades da CVS, a saber: “Desenho e Transição”, “Obter ou Construir”, “Entrega e Suporte” e/ou “Melhorar”. A execução dessas atividades gera “Produtos e Serviços” que, por fim, entregam “Valor” para o consumidor do serviço, que pode ser o próprio cliente, usuário(s) e/ou patrocinador(es).

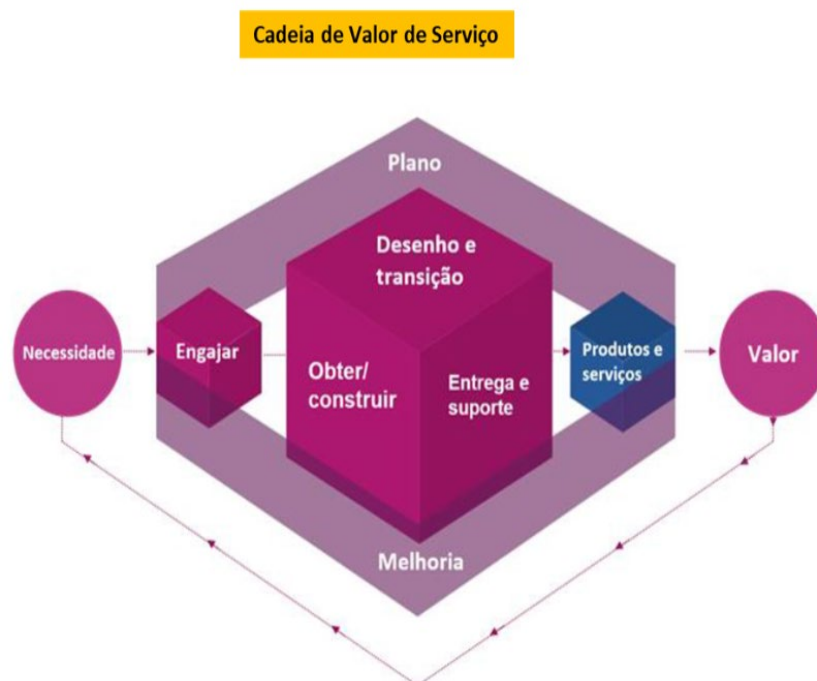


FIGURA 1. Figura baseada no material oficial da AXELOS ITIL®, apresentada no curso “ ITIL4 Fundamentos” da empresa ViaVirtual: <<https://viavirtual.net/courses/itil-r-4-fundamentos>>

### **Geração de Valor a partir da Oferta de Repositórios Institucionais de Dados de Pesquisa**

Nesta seção, são descritos os valores percebidos no cumprimento do Edital Público RNP-IBICT-CNPq para promoção da Ciência Aberta no Brasil. UFRA, UFC, UFG e CBPF configuram-se como os consumidores de valor, a partir do



recebimento de ações de transferência de conhecimento e capacitação para ofertarem seus repositórios institucionais de dados de pesquisa. O GT-RDP configura-se como o produtor de valores através das ações realizadas com as ICTs. RNP, IBICT e CNPq são os patrocinadores por meio dos quais as ações executadas são justificadas.

O projeto denominado “REDEDADOSABERTOS” foi liderado pelo GT-RDP Brasil através de reuniões de atendimento, ora coletivas, com todas as ICTs selecionadas, ora individuais, com apenas uma das ICTs selecionada. A coordenação acadêmica do GT-RDP Brasil manteve-se em constante alinhamento com a coordenação executiva do projeto, representada pela RNP, IBICT e CNPq, que também participava de algumas reuniões para acompanhamento das atividades e apresentação dos resultados.

A seguir, a instanciação de todos os elementos da Cadeia de Valor de Serviço é apresentada de forma a elucidar a geração de valor a partir da oferta de repositórios institucionais de dados de pesquisa.

Entrada

Necessidade

- Cumprimento do Edital Público RNP-IBICT-CNPq para promoção da Ciência Aberta no Brasil

Oportunidade/Demanda

- Execução do projeto denominado “REDEDADOSABERTOS”

### Ações da atividade “ENGAJAR”

- Aceitação pelas ICTs selecionadas para participarem do REDEDADOSABERTOS
- Participação das ICTs nas reuniões de atendimentos, ora semanais, ora quinzenais, do REDEDADOSABERTOS
- Aceitação da coordenação acadêmica pelo GT-RDP Brasil
- Aceitação da coordenação executiva da RNP junto com o IBICT
- Aceitação do CNPq como parceiro estratégico na oferta do CoNCiência (consórcio que estabelece bases de cooperação técnica e operacional, visando à promoção de atividades de incentivo à prática da Ciência Aberta [<https://www.gov.br/cnpq/pt-br/assuntos/noticias/cnpq-em-acao/cnpq-lancou-consorcio-conciencia-para-incentivar-a-pratica-da-ciencia-aberta-1>]) para ingresso das ICTs selecionadas

### Ações da atividade “PLANEJAR”

- Reuniões de alinhamento, ora semanais, ora quinzenais, da coordenação de P&D da RNP com a coordenação acadêmica do GT-RDP
- Reuniões de alinhamento, esporádicas, da coordenação executiva da RNP, IBICT e CNPq

### Ações da atividade “DESENHO e TRANSIÇÃO”

- Proposta de uma instalação customizada nacional de Dataverse para instanciação por cada ICT selecionada
- Proposta de organização por comunidades (produtor - consumidor), estudos, conjuntos de dados e seus respectivos metadados, ... para instanciação por cada ICT selecionada

### Ações da atividade “OBTER / CONSTRUIR”

- Construção do repositório institucional de dados de pesquisa por ICT selecionada, através da obtenção da versão mais atualizada do Dataverse e ingestão dos dados de pesquisa respectivos das ICTs.

Trabalho realizado pela equipe de tecnologia da informação de cada ICT orientado pela equipe do RDP.

- Instanciação de uma organização de comunidades (produtor - consumidor), estudos, conjuntos de dados e seus respectivos metadados, ... em cada repositório institucional de dados de pesquisa por ICT selecionada. Trabalho realizado pela equipe de ciência da informação de cada ICT orientado pela equipe do RDP.
- Obtenção e apresentação de exemplos de políticas e materiais de apoio para oferta, operação e sustentação de repositórios institucionais de dados de pesquisa. Trabalho realizado pela equipe do RDP.
- Construção de templates de políticas e materiais de apoio para oferta, operação e sustentação de repositórios institucionais de dados de pesquisa baseados nos exemplos apresentados pela equipe do RDP. Trabalho realizado em conjunto por todos os participantes do REDEDADOSABERTOS.
- Construção de políticas e materiais de apoio específicos para oferta, operação e sustentação dos repositórios institucionais de dados de pesquisa respectivos das ICTs selecionadas. Trabalho realizado por cada ICT selecionada e revisado pelo RDP.

#### Ações da atividade “ENTREGA e SUPORTE”

- Como a atividade de OBTER / CONSTRUIR foi feita de forma assistida pelo RDP com cada uma das ICTs, não houve a atividade de ENTREGA. A atividade de SUPORTE foi realizada pontualmente, em caso de dúvidas individuais, e, coletivamente, nas reuniões com todas as ICTs.

#### Ações da atividade “MELHORAR”

- A cada atendimento do REDEDADOSABERTOS era avaliado, na reunião de alinhamento entre coordenação de P&D da RNP e equipe do RDP, se havia como melhorar os atendimentos às ICTs. Em alguns momentos, essa avaliação levou a equipe do RDP a realizar atendimentos

individualizados para cada uma das ICTs, em outros momentos, foi percebido que o ganho do atendimento seria maior com o atendimento conjunto de todos no REDEDADOSABERTOS.

- Foi percebido que a execução de tarefas pelas ICTs e a abertura para cada uma delas apresentar seus resultados no atendimento conjunto do REDEDADOSABERTOS trouxe dinâmica para as reuniões de atendimentos e melhor resultados. Por exemplo, em uma reunião de atendimento do RDP foi apresentado um template de "Política de Acesso" para repositórios institucionais de dados de pesquisa e, em duas reuniões seguintes, dois pares de ICTs apresentaram suas respectivas instanciações de Políticas de Acesso. Situação semelhante também ocorreu com a pesquisa para abertura de dados de pesquisa nas ICTs.
- Foi percebido melhora quando as ICTs atuaram como protagonistas nas plenárias, acontecidas em cada 2 meses ao longo do REDEDADOSABERTOS. Os resultados individuais foram trabalhados e apresentados por cada uma das ICTs selecionadas. Com destaque para a plenária final, onde cada uma das ICTs apresentaram seus respectivos repositórios institucionais de dados de pesquisa funcionais como provas de conceito.

Saída

Produtos

- Provas de Conceitos de repositórios institucionais das ICTs do REDEDADOSABERTOS com seus respectivos dados de pesquisa.
- Vídeos curtos de capacitação disponibilizados na playlist do IBICT denominada "REDE DADOS ABERTOS Compartilhamento de Dados de Pesquisa em Repositórios", disponibilizada em: <<https://www.youtube.com/playlist?list=PLSk7nwl2wEh9CBEEqFRB6E5BqXiiC7HNP>>

## Serviços

- Consultoria especializada em P&D para abertura de dados de pesquisa em repositório institucional com experimentação em Dataverse.
- Orquestração para promover a troca de experiência e conhecimentos entre os participantes do REDEDADOSABERTOS liderada pela coordenação acadêmica do GT-RDP alinhada à coordenação executiva da RNP com IBICT e parceria com o CNPq.

## Valor

A saída de uma CVS é caracterizada por geração de valor. Valor é determinado por percepções do consumidor, o qual consome um produto e/ou serviço. Consumidores de serviço são caracterizados por: cliente, quem define e aceita o serviço; usuário, quem utiliza o serviço; e patrocinador, quem autoriza o orçamento para as custas do serviço.

Para as ICTs, o consumo do valor gerado foi relatado através da percepção da oferta de seus respectivos repositórios institucionais com seus dados de pesquisa instanciados.

Para os usuários, produtores e consumidores, o consumo do valor gerado foi relatado através da percepção de uso de seus respectivos repositórios institucionais de dados de pesquisa.

Para os patrocinadores, RNP, IBICT e CNPq, o consumo do valor gerado foi relatado pela criação de uma rede colaborativa para acesso aberto a dados de pesquisa, onde ações de promoção da Ciência Aberta no Brasil foram experimentadas nos repositórios institucionais de dados de pesquisa das ICTs.

## Conclusões

Esse artigo apresentou a experiência de geração de valor, a partir da oferta de repositórios institucionais de dados de pesquisa, vivenciada entre o período do segundo semestre do ano de 2021 e o primeiro semestre do ano de 2022 no

cumprimento do Edital Público RNP-IBICT-CNPq para promoção da Ciência Aberta no Brasil. As ICTs selecionadas, UFRA, UFG, UFC, além do CBPF, tiveram ações de capacitação lideradas pelo grupo de pesquisa do GT-RDP Brasil da UFRGS. A coordenação executiva das atividades foi da RNP e IBICT em parceria com o CNPq. Após o término do período previsto do edital, uma rede colaborativa embrionária para acesso aberto de dados de pesquisa foi composta, de forma voluntária e participativa, e que vem se integrando ao CoNCiência, Consórcio Nacional para Ciência Aberta no Brasil, liderado pelo CNPq.

## Bibliografia

- AXELOS. ITIL@ 4. <<https://www.axelos.com/certifications/itil-service-management>>
- BERTIN, P. R. B., FORTALEZA, J. M., SILVA, A. C., OKAWACHI, M. F. (2019). A parceria para governo aberto como plataforma para o avanço da Ciência Aberta no Brasil. *Transinformação*, 31: e190020. DOI: <[10.1590/2318-0889201931e190020](https://doi.org/10.1590/2318-0889201931e190020)>
- BRASIL. (2017). Acordo de cooperação técnica nº 2/2017 entre IBICT e RNP. *Diário Oficial da União*, 234, 14.
- CONTROLADORIA-GERAL DA UNIÃO (BRASIL). (2014). 4º Plano de Ação Brasileiro [Internet]. Brasília, DF. <<https://www.gov.br/cgu/pt-br/governo-aberto/a-ogp/planos-de-acao/4o-plano-de-acao-brasileiro/>>
- CONTROLADORIA-GERAL DA UNIÃO (BRASIL). (2018a). 4º Plano de Ação Brasileiro: monitoramento e execução [Internet]. Brasília, DF. <<https://www.gov.br/cgu/pt-br/governo-aberto/a-ogp/planos-de-acao/4o-plano-de-acao-brasileiro/4o-plano-de-acao-brasileiro-monitoramento-e-execucao>>
- CONTROLADORIA-GERAL DA UNIÃO (BRASIL). (2018b). Inovação e governo aberto na ciência: monitoramento e execução [Internet]. Brasília, DF, 29 Out 2018. <<https://www.gov.br/cgu/pt-br/governo-aberto/a-ogp/planos-de-acao/4o-plano-de-acao-brasileiro/compromisso-3-docs/inovacao-e-governo-aberto-na-ciencia-monitoramento-e-execucao>>
- KING, G. (2007). An introduction to the dataverse network as an infrastructure for data sharing. *Sociol Methodol Res*, 36, 173-199. <<https://j.mp/2owjuRr>>

- KING, G. Dataverse Project [Internet]. Cambridge, MA. Aplicativo. <<https://dataverse.org/>>
- RNP: ORGANIZAÇÃO SOCIAL DO MCTI. (2017a). RNP e Ibict lançam chamada para promover o acesso aberto a dados de pesquisa [Internet]. Brasília, DF, 15 Mar 2017. <<https://www.rnp.br/noticias/rnp-e-ibict-lancam-chamada-para-promover-o-acesso-aberto-dados-de-pesquisa>>
- RNP: ORGANIZAÇÃO SOCIAL DO MCTI. (2017b). RNP divulga proposta selecionada pela chamada de dados abertos para pesquisa [Internet]. Brasília, DF, 12 Maio 2017. <<https://www.rnp.br/noticias/rnp-divulga-proposta-selecionada-pela-chamada-de-dados-abertos-para-pesquisa>>
- RNP: ORGANIZAÇÃO SOCIAL DO MCTI. (2021). RNP, CNPq e IBICT divulgam resultado do edital para incubação de repositórios de dados de pesquisa [Internet]. Brasília, DF, 18 Ago 2021. <<https://www.rnp.br/noticias/rnp-cnpq-e-ibict-divulgam-resultado-do-edital-para-incubacao-de-repositorios-de-dados-de>>
- UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL. (2022). Centro de Documentação e Acervo Digital da Pesquisa Rede de Dados de Pesquisa Brasileira (RDP Brasil). RDPBrasil [Internet]. Porto Alegre. <<https://www.ufrgs.br/rdpbrasil/wp/>>

# Modelo dimensional para la medición de la producción académica

## **PABLO CÉSAR DE ALBUQUERQUE**

CESGI COMISIÓN DE INVESTIGACIONES CIENTÍFICAS

[pablo@sedici.unlp.edu.ar](mailto:pablo@sedici.unlp.edu.ar)

## **GONZALO LUJÁN VILLARREAL**

PREBI-SEDICI UNIVERSIDAD NACIONAL DE LA PLATA

CESGI COMISIÓN DE INVESTIGACIONES CIENTÍFICAS

[gonzalo@prebi.unlp.edu.ar](mailto:gonzalo@prebi.unlp.edu.ar)

## **MARISA RAQUEL DE GIUSTI**

PREBI-SEDICI UNIVERSIDAD NACIONAL DE LA PLATA

CESGI COMISIÓN DE INVESTIGACIONES CIENTÍFICAS

[marisa.degiusti@sedici.unlp.edu.ar](mailto:marisa.degiusti@sedici.unlp.edu.ar)

### **EJE TEMÁTICO**

Evaluación y métricas alternativas

### **RESUMEN**

Este artículo se centra en la problemática de la integración y análisis conjunto de la producción académica de las instituciones argentinas, que se encuentra dispersa en diversas fuentes de información, como repositorios digitales, portales de revistas y sistemas CRIS. Se propone un modelo dimensional basado en Kimball, independientemente de las fuentes de datos utilizadas, con el fin de tener un modelo que permita el análisis de datos provenientes de múltiples fuentes, a lo largo del tiempo.

### **PALABRAS CLAVE**

Data warehouse; repositorios institucionales; business intelligence.



## Introducción

La gestión eficiente de grandes volúmenes de datos académicos es un desafío crucial en el contexto actual de la investigación y la educación superior. La creciente producción académica y la diversidad de fuentes y formatos de datos ha llevado a la necesidad de desarrollar enfoques efectivos para organizar, integrar y analizar esta información de manera sistemática (Xia et al., 2017). En el ámbito académico, la generación de datos involucra una amplia gama de entidades, incluyendo instituciones, colaboradores y publicaciones científicas. La cantidad y complejidad de datos vinculados a estas entidades requiere una gestión eficiente que permita su organización, consulta y análisis de manera efectiva.

Los desafíos asociados con la gestión de datos académicos son diversos. La diversidad de fuentes y formatos de datos dificulta la integración y el análisis holístico de la información. Además, la falta de estándares para la representación de entidades académicas, la necesidad de desambiguación de nombres y la dificultad para relacionar datos de diferentes fuentes agregan complejidad al proceso de gestión de datos académicos (De Albuquerque et al., 2021). Nuestro objetivo principal en este trabajo es diseñar un modelo dimensional que permita una gestión eficiente de datos académicos a lo largo del tiempo. Este enfoque se basa en la estructuración de la información relevante sobre instituciones, colaboradores y publicaciones en tablas estructuradas, aprovechando identificadores persistentes y vocabularios controlados para mejorar la integración y búsqueda de datos. Al utilizar este enfoque, es factible diseñar a futuro un esquema de Data Warehouse que integre las dimensiones y los hechos propuestos en este trabajo junto con procesos de ETL que permitan integrar, organizar y estructurar grandes volúmenes de datos, que luego podrán ser consultados y explotados por medio de herramientas de Business Intelligence como Power BI, Tableau o Google Looker Studio (ex Data Studio).

## **Metodología**

En este artículo, nos basaremos en el Proceso de Diseño Dimensional de Cuatro Pasos propuesto por Ralph Kimball (Kimball & Ross, 2013). Este proceso está compuesto por cuatro etapas que son:

- 1) Seleccionar el proceso de negocio
- 2) Declarar la granularidad
- 3) Identificar las dimensiones
- 4) Identificar los hechos

### ***Selección del proceso de negocio***

Los procesos de negocio son las actividades operativas realizadas por una organización, como por ejemplo tomar un pedido, procesar un reclamo o registrar estudiantes para una clase. En el enfoque dimensional de Kimball, la elección del proceso de negocio es importante porque define un objetivo de diseño específico y permite declarar la granularidad, las dimensiones y los hechos correspondientes. La mayoría de las tablas de hechos se centran en los resultados de un solo proceso de negocio.

Además, cada proceso de negocio se correlaciona con una fila en la matriz de buses del almacén de datos empresariales. La matriz de buses es una herramienta utilizada para organizar y relacionar las diferentes dimensiones y procesos de negocio en un diseño dimensional coherente.

### ***Declarar la granularidad***

La granularidad se refiere al nivel de detalle con el que los datos son capturados en un modelo dimensional estableciendo qué representa exactamente una sola fila de la tabla de hechos. Es necesario definir la granularidad antes de elegir las dimensiones o los hechos, ya que todas las dimensiones o hechos potenciales deben ser consistentes con la granularidad declarada.

Kimball menciona el concepto de "atomic grain" o granularidad atómica. Esto se refiere al nivel más bajo en el que se capturan los datos en un proceso de negocio específico. Se recomienda enfocarse en los datos capturados a nivel atómico porque son más flexibles para responder a consultas impredecibles de los usuarios.

En un enfoque de granularidad atómica, los datos se representan en su forma más desglosada, capturando cada elemento individual. Por otro lado, la granularidad resumida implica la agregación o consolidación de datos en niveles más altos de abstracción, lo que implica una reducción en la especificidad de los detalles. Al declarar una granularidad resumida, es posible que solo se respondan las preguntas más frecuentes y se pierda la oportunidad de realizar un análisis más profundo. Si bien las granularidades resumidas o agregadas son útiles para optimizar el rendimiento, es importante tener en cuenta que pueden limitar la exploración de datos en detalle y restringir el alcance de las respuestas obtenidas. Esto implica que, al utilizar una granularidad resumida, es fundamental considerar cuidadosamente las preguntas comunes del negocio que se desean responder, ya que los niveles más detallados de los datos pueden no estar disponibles en la visualización o análisis.

### ***Definir dimensiones para describir el contexto***

Las dimensiones proporcionan el contexto de "quién, qué, dónde, cuándo, por qué y cómo" que rodea un evento del proceso de negocio. Las tablas de dimensiones contienen atributos descriptivos utilizados por las aplicaciones de inteligencia de negocios (BI) para filtrar y agrupar los hechos. Teniendo en cuenta claramente la granularidad de la tabla de hechos, es posible identificar todas las dimensiones posibles.

Cuando una dimensión está asociada con una fila de hecho, preferiblemente debería tener un único valor. Las tablas de dimensiones contienen las etiquetas

descriptivas que permiten aprovechar el sistema de DW/BI para el análisis de datos.

### ***Definir hechos para medir***

Los hechos son las mediciones que resultan de un evento del proceso de negocio y, en su mayoría, son valores numéricos. Una fila de la tabla de hechos tiene una relación uno a uno con un evento de medición tal como se describe en la granularidad de la tabla de hechos. Por lo tanto, una tabla de hechos corresponde a un evento físico observable y no a las necesidades de un informe en particular.

### ***Extensiones y adaptaciones posibles en modelos dimensionales ante cambios en los datos***

Los modelos dimensionales son resistentes a los cambios en las relaciones de datos. Estas extensiones se pueden implementar sin alterar las consultas o aplicaciones existentes de inteligencia de negocios (BI) y sin afectar los resultados de las consultas.

A continuación, se presentan los tipos de cambios que se pueden realizar en un modelo dimensional:

- Se pueden agregar nuevos hechos que sean consistentes con la granularidad existente de una tabla de hechos mediante la creación de nuevas columnas.
- Se pueden agregar nuevas dimensiones a una tabla de hechos existente mediante la creación de nuevas columnas de clave externa, siempre y cuando no alteren la granularidad de la tabla de hechos.
- Se pueden agregar nuevos atributos a una tabla de dimensiones existente mediante la creación de nuevas columnas.
- Se puede aumentar la granularidad de una tabla de hechos agregando atributos a una tabla de dimensiones existente y luego redefiniendo la tabla de hechos a una granularidad más baja. Es importante asegurarse

de preservar los nombres de columna existentes en las tablas de hechos y dimensiones.

Estas extensiones permiten adaptar y ampliar los modelos dimensionales sin afectar las consultas y aplicaciones existentes.

## **Propuesta**

### ***Selección del proceso de negocio***

En el contexto de este estudio, el proceso de negocio seleccionado se centra en la publicación de recursos académicos generados por una institución científico-académica. El objetivo principal es medir y analizar la producción académica de la institución, estableciendo las bases para futuras evaluaciones del rendimiento y el impacto.

El proceso de publicación abarca una variedad de sistemas y plataformas utilizadas para difundir los recursos académicos. Se consideran fuentes de datos relevantes, como los repositorios institucionales, las revistas científicas, las conferencias, las bases de datos especializadas y los buscadores web, entre otros. Se prioriza el uso de servicios open source y fuentes de datos abiertas para promover la adopción de este modelo en diferentes instituciones científicas y académicas de Argentina.

El período de análisis no presenta limitaciones específicas, y se busca incluir la mayor cantidad de publicaciones posible. Esto permitirá obtener una visión global de la producción académica de la institución a lo largo del tiempo.

Entre los atributos a recopilar para cada publicación académica se encuentran:

- Autor: Identificación del autor o autores responsables del recurso académico.
- Institución: Afiliación institucional del autor o autores.

- Fecha de publicación: Información sobre el momento en que se realizó la publicación.
- Derechos de acceso (*access rights*): Licencia de uso asociada al recurso académico.
- Tipo de recurso: Clasificación del recurso académico, como artículos de investigación, libros, contribuciones a conferencias, entre otros.

### **Declaración de granularidad**

El nivel de detalle deseado para cada evento de medición es a nivel de cada recurso académico individual. Esto implica capturar información específica sobre cada recurso, que puede estar compuesto por archivos de texto, licencias, imágenes, entre otros. Además, se busca representar a cada autor y las relaciones de dependencia entre ellos, para tener en cuenta casos en los que un autor esté afiliado a múltiples instituciones. En este caso, la granularidad está definida por la publicación de recursos identificables en la web a través de identificadores persistentes, como DOI o Handle en el caso de publicaciones, ORCID para los autores y ROR para las instituciones. Estos identificadores permiten la intersección de datos provenientes de diferentes fuentes, facilitando así la medición y el análisis (Sompel et al., 2014).

La información sobre los eventos de producción académica se captura y registra a través de sistemas y bases de datos específicas. En particular, en una primera instancia, se pueden utilizar los repositorios institucionales propios de cada institución, donde se define qué es relevante para cada una de estas. Estos repositorios cuentan con personal dedicado a revisar, corregir y normalizar los recursos provenientes de otras fuentes de datos. Esto garantiza un nivel mínimo de calidad de los datos para su posterior procesamiento. Además, se aprovecha la interoperabilidad y las directrices definidas a nivel nacional para asegurar la disponibilidad de los metadatos a través de OAI-PMH. Una vez explotados los repositorios propios, pueden considerarse nodos nacionales o sistemas agregadores, en busca de recursos de autores propios de la institución cargados en otros repositorios, ya sea en casos de múltiple

dependencia o en coautoría con investigadores de otras instituciones. En una tercera instancia, pueden considerarse fuentes externas como Unpaywall, Crossref o Scopus, prestando especial atención a la disponibilidad de herramientas de consulta y acceso a datos (OAI-PMH, API REST, OpenSearch, etc).

Existen restricciones en cuanto a la disponibilidad de datos a nivel de granularidad deseado. En principio se busca utilizar fuentes de datos abiertas siempre que sea posible, y en el caso particular de Argentina, los repositorios institucionales que forman parte del SNRD, ofrecen la exposición de metadatos de su producción académica (Directrices SNRD, 2015). Existen modelos similares en otros países de la región que también pueden aprovecharse, como por ejemplo el SIC en Chile, KIMUK en Costa Rica, REMERI en México o SILO en Uruguay. E incluso es viable considerar a LA Referencia como fuente de acceso a recursos provenientes de 12 nodos nacionales (LA Referencia - Nodos, s. f.). Sin embargo, se reconoce la utilidad de fuentes de datos cerradas propias de cada institución, que agilicen la ingesta y el procesamiento de los datos académicos. Aunque es importante tener en cuenta que estas fuentes cerradas pueden requerir adaptaciones para su reutilización en otros contextos.

### ***Identificación de dimensiones***

Antes de adentrarnos en la identificación de las dimensiones relevantes para medir la producción académica de la institución, es importante destacar que en nuestro modelo hemos adoptado una convención de nombres para las tablas. Todas las dimensiones se representan en la base de datos con el prefijo 'dim\_' seguido de un nombre descriptivo. Esta convención nos permite identificar fácilmente las tablas relacionadas con las dimensiones en el esquema de la base de datos. Además, hemos optado por utilizar nombres en inglés para mantener coherencia con las buenas prácticas de modelado dimensional y la nomenclatura ampliamente utilizada en el campo de la ciencia de datos.

Las dimensiones identificadas son:

- Institución
- Colaborador
- Disponibilidad o nivel acceso
- Tipo de recurso

Asimismo, en todas las dimensiones, se registran datos adicionales como la fecha de creación del registro ('created') en la fuente de datos, la fecha de modificación ('changed') una vez incorporado al modelo, y un valor arbitrario que refleja el grado de confianza ('confidence') basado en el origen y calidad de los datos. Además, en este modelo se han incluido tablas auxiliares con el sufijo '\_helper' que desempeñan un papel fundamental en el proceso de ingesta de datos, desambiguación y normalización. Estas tablas auxiliares contienen atributos adicionales que describen a cada registro en mayor detalle, como el nombre y apellido de un colaborador o el nombre y sigla de una institución.

Además de los atributos descriptivos, las tablas auxiliares también incluyen los campos 'source' y 'source\_id', que indican el nombre de la fuente de datos de origen (en un formato legible para el usuario) y el identificador correspondiente en dicha fuente de datos. Estas tablas auxiliares desempeñan un papel crítico en garantizar la calidad y coherencia de los datos de las dimensiones en el modelo dimensional. Al incluir información adicional en estas tablas, se simplifica y agiliza el proceso de carga y transformación de los datos, brindando un soporte fundamental para la integración de datos provenientes de diferentes fuentes. Además, al separar esta información auxiliar de la tabla principal, se facilita la visualización y la interpretación directa por parte del usuario final, centrándose en los atributos relevantes para el análisis de la producción académica de la institución.

### ***Dimensión Institución***

La dimensión 'Institución' se encuentra representada en la tabla 'dim\_institution'. Esta dimensión es fundamental para capturar información



descriptiva sobre las instituciones académicas. En dicha tabla, se almacenan atributos como el nombre, sigla y tipo de institución. Además, cabe destacar que la tabla 'dim\_institution' cuenta con seis atributos que hacen referencia a identificadores de instituciones que son consideradas como padres para cada registro. Esta relación jerárquica permite establecer la estructura organizativa de las instituciones. Asimismo, se han incluido otros seis atributos que almacenan los nombres normalizados de las instituciones, facilitando su interpretación por parte de los usuarios finales. Un campo adicional, denominado 'doble\_dependencia', indica si una institución tiene más de un padre directo, brindando información relevante sobre las relaciones de dependencia entre las instituciones.

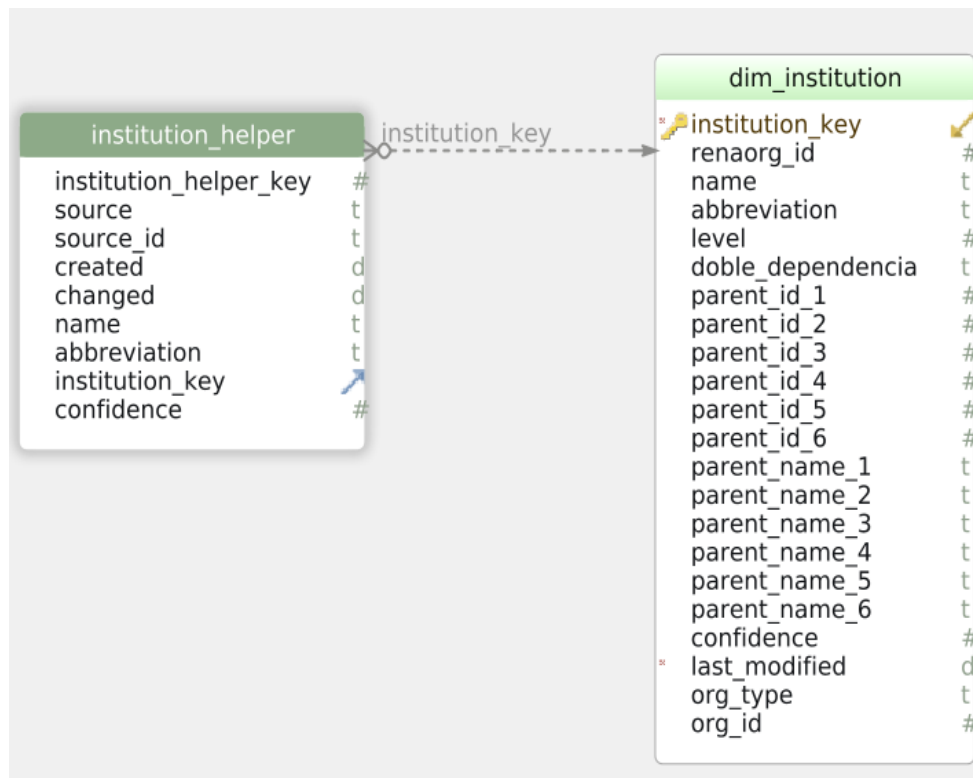


IMAGEN 1. Esquema de la de la dimensión 'Institución'

La dimensión 'Institución' utiliza el Registro de Organizaciones de Investigación (ROR) como la clave principal o business key para la identificación única de las instituciones. ROR es un registro a nivel mundial,

impulsado por la comunidad, que proporciona identificadores persistentes y abiertos para las organizaciones de investigación. La elección de ROR como fuente de identificación se basa en su amplia adopción en la comunidad académica y científica, así como en su disponibilidad como recurso de acceso abierto (French, 2023). Esto facilita la reutilización de los datos y garantiza la integridad de las referencias a las instituciones en el modelo dimensional. Asimismo, para el caso de instituciones Argentinas, es posible tomar el identificador nacional RENAORG como *business key* complementaria.

Además, se ha creado una tabla auxiliar denominada 'institution\_helper' para almacenar datos adicionales que son relevantes en el proceso de ingesta de datos, desambiguación y normalización, pero que no son necesarios para la visualización o interpretación directa por parte del usuario final.

### **Dimensión Colaborador**

La dimensión 'Colaborador' representa a las personas que han contribuido en la producción académica, ya sea como autores principales, coautores, directores de tesis de postgrado, traductores u otros roles relevantes en el proceso. Esta dimensión registra datos en la tabla "dim\_contributor" y almacena información sobre los colaboradores involucrados en la producción académica, permitiendo un análisis detallado de su participación. Tanto los autores principales como los colaboradores desempeñan un papel fundamental en la generación del conocimiento y es importante incluirlos en el modelo dimensional para comprender completamente la producción académica de la institución.

La *business key* utilizada para identificar de manera única a cada colaborador en la dimensión 'Colaborador' es el identificador ORCID (Open Researcher and Contributor ID). ORCID es un sistema globalmente reconocido y adoptado por la comunidad académica que proporciona identificadores persistentes y únicos para investigadores y colaboradores. La elección de ORCID como *business key* se debe a su amplia adopción y a su disponibilidad

en acceso abierto, lo que facilita la integración de datos y promueve la reutilización de la información académica (Petro, 2020). Mediante el uso de ORCID, se asegura la integridad y la trazabilidad de los colaboradores en el análisis de la producción académica de la institución.

La dimensión 'Colaborador' registra datos como el nombre, apellido y número de teléfono, así como identificadores ampliamente adoptados en otros servicios como Google Scholar, Publons y Scopus. Si bien existe una tabla auxiliar asociada a la dimensión para referenciar nuevos orígenes de datos, los identificadores de Google Scholar, Scopus, Publons, y las páginas web personales de los colaboradores se almacenan directamente en la tabla 'dim\_contributor'. Esta elección busca aprovechar que las personas ya están familiarizados con el uso de estos identificadores y por lo tanto simplificarán las tareas de generación de informes y reportes a partir de los datos alojados.

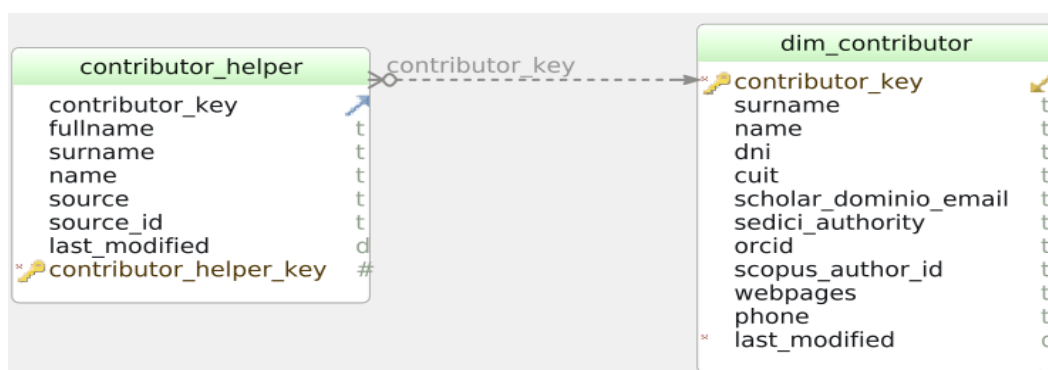


IMAGEN 2. Esquema de la dimensión 'Colaborador'

### ***Rol de colaborador en una publicación***

Una misma persona puede tener diferentes roles en distintas publicaciones. Por ejemplo, puede ser autor de un artículo, director de una tesis y compilador de un libro. Para representar el rol desempeñado por una persona en una publicación se utiliza la tabla 'bridge\_contributor\_publication'. Además de contener las claves foráneas que hacen referencia al colaborador y su

respectiva publicación, esta tabla incluye un atributo denominado 'role', el cual almacena inicialmente el rol en forma de una cadena de texto.

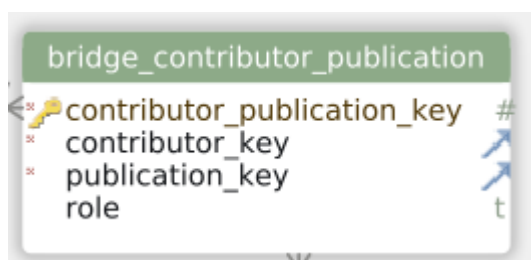


IMAGEN 3. Esquema del rol de colaborador en una publicación

### ***Dimensión de Disponibilidad o derechos de acceso***

La dimensión de 'Disponibilidad' se encuentra representada en la tabla 'dim\_access\_right'. La misma se utiliza para modelar los diferentes niveles de acceso que pueden aplicarse a un recurso académico. Para establecer la business key de esta dimensión, se ha tomado como referencia el vocabulario controlado definido por COAR (COAR Controlled Vocabularies Interest Group, 2022). En dicha tabla se registran valores como embargoed access (acceso con embargo), metadata only access (acceso solo a metadatos), open access (acceso abierto) y restricted access (acceso restringido). Estos atributos permiten representar de manera precisa y estandarizada las políticas de disponibilidad de los recursos académicos en la web.

Esta dimensión está compuesta por varios atributos que proporcionan información adicional sobre los niveles de acceso de un recurso académico. Entre estos atributos se encuentran 'label', que contiene el nombre legible en inglés para el usuario final, y 'label\_es', que corresponde a su traducción al español para una mejor comprensión. Además, se incluye el atributo 'uri', que almacena la URI del recurso en la web, permitiendo su fácil acceso.

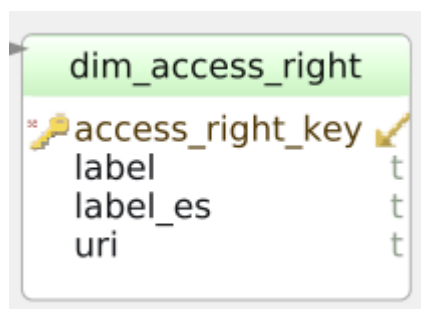


IMAGEN 4. Esquema de la dimensión de 'Disponibilidad'

### ***Dimensión de Tipo de Recurso***

La tabla 'dim\_resource\_type' es utilizada para representar los diferentes tipos de recursos que se generan como parte de la producción académica de una institución. En esta dimensión, se almacenan los datos relacionados con los tipos de recursos, siguiendo un modelo basado en el vocabulario controlado de COAR (COAR Controlled Vocabularies Interest Group, 2022).

La dimensión 'dim\_resource\_type' cuenta con varios atributos que enriquecen su contenido. Entre ellos se encuentran 'label' y 'label\_es', que contienen las etiquetas en inglés y español respectivamente, para una mejor comprensión por parte del usuario final. También se incluye 'definition\_en', que proporciona una descripción del alcance del tipo de recurso en inglés.

Además, se registra la URI del recurso en la web, permitiendo su acceso directo. Asimismo, se maneja una estructura jerárquica en este vocabulario controlado, por lo que se incluye la URI del padre, para reflejar las relaciones entre los diferentes tipos y subtipos de documentos.

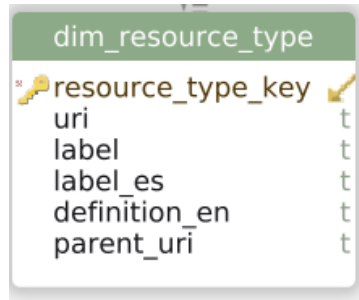


IMAGEN 5. Esquema de la dimensión de 'Tipo de Recurso'

## Identificación de hechos

### *Hecho publicación*

El hecho de publicación se registra en la tabla "fact\_publication" y contiene información detallada sobre la publicación de un recurso académico específico. Para garantizar la unicidad de cada publicación, se utiliza una business key preferentemente basada en el identificador persistente conocido como DOI (Digital Object Identifier), el cual es mantenido por Crossref (Hendricks, 2021). En caso de que una publicación no disponga de un DOI, se utiliza el atributo "authority" para almacenar otro identificador válido, como por ejemplo "handle", ampliamente utilizado en plataformas como DSpace (Donohue, 2021).

Con el fin de facilitar la integración de nuevas publicaciones y mejorar la interoperabilidad, se ha desarrollado una tabla auxiliar llamada "helper" que contiene otros identificadores relacionados con las publicaciones. Estos identificadores adicionales se almacenan para asistir en el proceso de integración de datos de nuevas publicaciones, siguiendo la misma estrategia aplicada en las dimensiones de institución y colaborador.

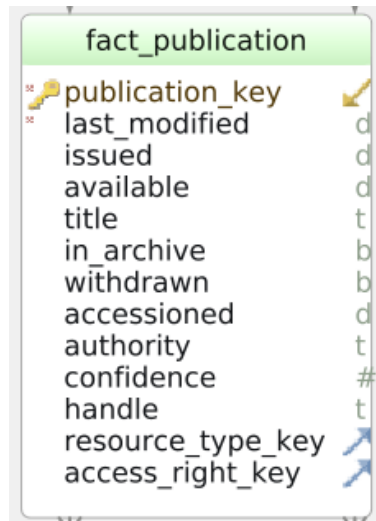


IMAGEN 6. Esquema del hecho 'Publicación'

### **Hecho filiación**

La tabla de hechos de 'Filiación', representado por la tabla "fact\_affiliation", busca capturar la relación existente entre un colaborador (contributor) y una institución en un período de tiempo específico. Para lograr esto, se utilizan claves foráneas que hacen referencia a las dimensiones de colaborador (contributor) e institución (institution), junto con los atributos de fecha de inicio y fecha de finalización para indicar la validez de esta relación en el tiempo.

En este caso, no se han identificado *business keys* específicas que permitan identificar la filiación de una persona de manera única. Sin embargo, se puede utilizar la información de las publicaciones asociadas a un colaborador para calcular los valores de inicio y fin de la filiación. Por ejemplo, si una persona publica un artículo donde establece su afiliación con una institución X en el año 2018, podemos inferir que esa filiación existió al menos a partir de dicho año; si luego se procesa una publicación del mismo autor vinculada a la misma institución con año 2016, podrá registrarse que la persona ya pertenecía a la institución desde al menos 2016.

De esta manera, la tabla de hechos "fact\_affiliation" proporciona una representación precisa de la relación entre colaboradores e instituciones en un

contexto temporal, permitiendo un análisis más completo de la producción académica y las colaboraciones entre diferentes actores. Este modelo también resulta útil para representar la pertenencia a múltiples instituciones de una misma persona en un momento dado.



IMAGEN 7. Esquema de 'Filiación'

## Conclusión

En este trabajo, hemos abordado el desafío de diseñar un esquema dimensional para gestionar de manera eficiente un gran volumen de datos académicos a lo largo del tiempo. Nuestro enfoque se ha centrado en la representación estructurada de información relevante sobre instituciones, colaboradores y publicaciones en un modelo dimensional. Mediante la implementación de tablas como 'dim\_institution', 'dim\_contributor' y 'fact\_publication' en un sistema de bases de datos relacional, hemos logrado representar de manera precisa la información clave relacionada con estas entidades.

Nuestra contribución a la gestión de datos académicos radica en la propuesta de un modelo dimensional que mejora la eficiencia en la consulta y organización de la información. Utilizando identificadores persistentes ampliamente adoptados como DOI y ORCID, y aprovechando vocabularios controlados como los proporcionados por COAR, hemos facilitado la integración y búsqueda de datos académicos.



Al basarnos en la metodología propuesta por Kimball, nuestro modelo dimensional ofrece beneficios significativos en términos de estructura de datos, precisión en la identificación de instituciones y colaboradores, y eficiencia en el análisis de publicaciones académicas. Además, hemos destacado la importancia de considerar tablas auxiliares, como 'institution\_helper' y 'contributor\_helper', para apoyar tareas de ingesta de datos y desambiguación.

Como trabajo futuro, planeamos expandir este modelo para incorporar los hechos y dimensiones adecuados que permitan medir no solo la producción científica, sino también el impacto que esta genera, de acuerdo a los intereses de cada institución. Sumado a esto, como se mencionó previamente, este proyecto requiere una adecuada selección de fuentes de datos, considerando especialmente el nivel de confianza o la calidad de los datos que ofrece, la definición de patrones de ingesta y procesamiento de datos (ETL, ELT o ETLT) (Gupta et al., 2020) y la generación de una infraestructura capaz de ejecutar estos procesos, almacenar los datos de manera eficiente y generar reportes para usuarios finales.

## Bibliografía

- COAR CONTROLLED VOCABULARIES INTEREST GROUP. (2022, septiembre 29). Controlled Vocabularies for Repositories: Access Rights. <[https://vocabularies.coar-repositories.org/access\\_rights/](https://vocabularies.coar-repositories.org/access_rights/)>
- COAR CONTROLLED VOCABULARIES INTEREST GROUP. (2022, septiembre 29). Controlled Vocabularies for Repositories: Resource Types. <[https://vocabularies.coar-repositories.org/resource\\_types/](https://vocabularies.coar-repositories.org/resource_types/)>
- DE ALBUQUERQUE, P. C., VILLARREAL, G. L., & DE GIUSTI, M. R. (2021). Proposal of a Data Warehouse for Scholarly Institutions built on Institutional Repositories. IX Jornadas de Cloud Computing, Big Data & Emerging Topics (Modalidad virtual, 22 al 25 de junio de 2021). <<https://sedici.unlp.edu.ar/handle/10915/125161>>
- DIRECTRICES PARA PROVEEDORES DE CONTENIDO DEL SISTEMA NACIONAL DE REPOSITARIOS DIGITALES. (2015). [ebook] Buenos Aires: Sistema

Nacional de Repositorios Digitales. Ministerio de Ciencia, Tecnología e Innovación Productiva.  
<[https://repositoriosdigitales.mincyt.gob.ar/files/Directrices\\_SNRD\\_2015.pdf](https://repositoriosdigitales.mincyt.gob.ar/files/Directrices_SNRD_2015.pdf)>

DONOHUE, T. (2021). Handle.Net Registry Support—DSpace 7.x Documentation—LYRISIS Wiki.  
<<https://wiki.lyrasis.org/display/DSDOC7x/Handle.Net+Registry+Support>>

FRENCH, A. (2023). Case-studies. Research Organization Registry (ROR).  
<<https://ror.org/categories/case-studies/>>

GUPTA, A., SAHAYADHAS, A., & GUPTA, V. (2020). Proposed Techniques to Design Speed Efficient Data Warehouse Architecture for Fastening Knowledge Discovery Process. 2020 IEEE Third International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), 200-201. <<https://doi.org/10.1109/AIKE48582.2020.00039>>

HENDRICKS, G. (2021, diciembre 11). The research nexus [Website]. Crossref.  
<<https://www.crossref.org/documentation/research-nexus/>>

KIMBALL, R., & ROSS, M. (2013). The Data Warehouse Toolkit, 3rd Edition. Kimball Group. <<https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/books/data-warehouse-dw-toolkit/>>

LA REFERENCIA—NODOS. (s. f.). <<https://www.lareferencia.info/es/nodos>>

PETRO, J. (2020, octubre 20). 10M ORCID iDs! 10M ORCID IDs!  
<<https://info.orcid.org/10m-orcid-ids/>>

SOMPEL, H. V. DE, SANDERSON, R., SHANKAR, H., & KLEIN, M. (2014). Persistent Identifiers for Scholarly Assets and the Web: The Need for an Unambiguous Mapping. *International Journal of Digital Curation*, 9(1), Art. 1. DOI: <<https://doi.org/10.2218/ijdc.v9i1.320>>

XIA, F., WANG, W., BEKELE, T. M., & LIU, H. (2017). Big Scholarly Data: A Survey. *IEEE Transactions on Big Data*, 3(1), 18-35. DOI: <<https://doi.org/10.1109/TBDATA.2016.2641460>>

# Interoperabilidad automática gracias al uso de PIDs - ORCID

**ANA CARDOSO**

ORCID

[p.cardoso@orcid.org](mailto:p.cardoso@orcid.org)

## EJE TEMÁTICO

Infraestructura tecnológica

## RESUMEN

Uno de los conceptos clave de ciencia abierta es compartir, pero esto no es posible sin una infraestructura de investigación abierta. En la época actual los resultados y modelos de publicación y diseminación de la ciencia crecen con rapidez y la ciencia se vuelve cada vez más interconectada y global. Investigadores, bibliotecarios, instituciones de investigación, entes de financiamiento y gubernamentales deben hacer frente al desafío, deben rastrear y demostrar el impacto de las contribuciones de investigación. En la práctica, los identificadores persistentes (PIDs) ayudan actuando como coordenadas, indicando la proveniencia y las fuentes de información y mostrando las conexiones entre personas, instituciones y contribuciones. Individualmente, cada uno de estos identificadores es útil, pero su valor se potencia cuando se utilizan en los flujos de trabajo digitales, para crear y compartir fácilmente conexiones de confianza entre ellos. La infraestructura interoperable de ORCID puede ayudar a acelerar el descubrimiento del conocimiento y aumentar la transparencia y reproducibilidad de la investigación al fomentar los principios de datos FAIR (Findability, Accessibility, Interoperability, and Reusability) y las prácticas de Ciencia Abierta a través de identificadores persistentes y datos estandarizados y de acceso abierto. Los PIDs y sus metadatos asociados se pueden considerar los bloques edificantes de la infraestructura de investigación.

## PALABRAS CLAVES

Interoperabilidad; interoperability; identificadores persistentes; persistent identifiers; infraestructuras abiertas; open infrastructures.

# Taxonomía de la ciencia abierta: revisada y ampliada

## **LÚCIA DA SILVEIRA**

UNIVERSIDADE FEDERAL DE SANTA CATARINA (UFSC)

## **REMEDIOS MELERO**

INSTITUTO DE AGROQUÍMICA Y TECNOLOGÍA DE ALIMENTOS

## **NIVALDO CALIXTO RIBEIRO**

UNIVERSIDADE FEDERAL DE MINAS GERAIS

## **ANDREA MORA-CAMPOS**

UNIVERSIDAD NACIONAL

## **PRISCILA MACHADO BORGES SENA**

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

## **FABIANO COUTO CORRÊA DA SILVA**

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

## **JULIANA FACHIN**

GRUPO DE INVESTIGAÇÃO CIÊNCIA ABERTA LATAM

## **DANIEL FERNANDO PIRAQUIVE-PIRAQUIVE**

UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS

## **JORGE POLANCO CORTÉS**

UNIVERSIDAD DE COSTA RICA

## **RONALDO FERREIRA ARAÚJO**

UNIVERSIDADE FEDERAL DE ALAGOAS

## **ALEJANDRO URIBE TIRADO**

UNIVERSIDAD DE ANTIOQUIA

## **JULIO SANTILLÁN-ALDANA**

UNIVERSIDAD TECNOLÓGICA DEL PERÚ

## **ANDRÉS MAURICIO ENCISO BETANCOURT**

UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS

### **EJE TEMÁTICO**

Comunicación académica y científica en abierto

### **RESUMEN**

**Objetivo:** revisar las terminologías y aplicaciones de la taxonomía de la Ciencia Abierta, para construir una versión más completa, que represente el conocimiento en torno al tema, de acuerdo con el escenario actual de la comunicación científica y las recomendaciones de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (Unesco). **Método:** se trata de una investigación exploratoria con un enfoque deductivo. El primer paso fue la revisión de las taxonomías, con las 12 personas dedicadas a la investigación, que se reunieron semanalmente, para las discusiones conceptuales y epistemológicas relacionadas con la Ciencia Abierta y las definiciones metodológicas y de procedimiento para la realización del estudio. **Resultados:** como resultado de los análisis, se elaboró una taxonomía para ser evaluada por las personas expertas. Para ello, se envió a 68 personas expertas un cuestionario con preguntas abiertas sobre cada eje principal de la taxonomía. Se obtuvieron 21 respuestas que cooperaron con la modelización y exposición de los términos de la nueva propuesta de taxonomía. La nueva taxonomía tiene 10 componentes de nivel principal y un total de 96 etiquetas. **Conclusiones:** la percepción de los expertos puso de manifiesto un panorama congruente con las recomendaciones de la Unesco y el escenario actual de la Ciencia Abierta.

**PALABRAS CLAVES**

Ciencia Abierta; taxonomía; comunicación científica; representación del conocimiento; Unesco. Open Science; taxonomy; scientific communication; representation of knowledge; Unesco.



---

Pósters

---

# Rede Sudeste de Repositórios Digitais: a importância do trabalho colaborativo

**CLAUDETE FERNANDES DE QUEIROZ**

FUNDAÇÃO OSWALDO CRUZ

INSTITUTO DE COMUNICAÇÃO E INFORMAÇÃO CIENTÍFICA E TECNOLÓGICA EM SAÚDE

[claudete.queiroz@iciict.fiocruz.br](mailto:claudete.queiroz@iciict.fiocruz.br)

## RESUMO

Os desafios relacionados com a criação, otimização e sustentabilidade dos repositórios e a importância do trabalho cooperativo para o enfrentamento e superação de barreiras, bem como a necessidade de alinhamento das políticas de acesso aberto no Brasil e no âmbito internacional, impulsionou a criação da Rede Sudeste de Repositórios Digitais, que ocorreu em 3 de outubro de 2017 no Rio de Janeiro, através da assinatura da “Carta do Rio”, em que subscreveram 15 instituições de ensino e pesquisa. A Rede Sudeste tem realizado diversas reuniões para a discussão de diferentes questões que permitiram estabelecer um vínculo entre os profissionais envolvidos, caracterizando assim, a troca de experiências, os valores institucionais e uma maior percepção da importância da salvaguarda da produção científica dos repositórios envolvidos.



## Rede Sudeste de Repositórios Digitais: a importância do trabalho colaborativo

**Claudete Fernandes de Queiroz**

Fundação Oswaldo Cruz, Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Rio de Janeiro, RJ, Brasil, claudete.queiroz@icict.fiocruz.br

### Rede Sudeste de Repositório Digitais

Os desafios relacionados com a criação, otimização e sustentabilidade dos repositórios e a importância do trabalho cooperativo para o enfrentamento e superação de barreiras, bem como a necessidade de alinhamento das políticas de acesso aberto no Brasil e no âmbito internacional, impulsionou a criação da Rede Sudeste de Repositórios Digitais, que ocorreu em 3 de outubro de 2017, no Rio de Janeiro, através da assinatura da "Carta do Rio", em que subscreveram 15 instituições de ensino e pesquisa.

A Rede Sudeste tem realizado diversas reuniões para a discussão de diferentes questões que permitiram estabelecer um vínculo entre os profissionais envolvidos, caracterizando assim, a troca de experiências, os valores institucionais e uma maior percepção da importância da salvaguarda da produção científica dos repositórios envolvidos (figura 1).



Figura 1 - Reuniões realizadas 2017-2022

A Rede tem efetuado diversas ações visando a melhoria e a qualidade dos seus repositórios digitais, corroborando com os pressupostos do Movimento de Acesso Aberto que tem como objetivo "promover o acesso livre e irrestrito à literatura científica e acadêmica, favorecendo o aumento do impacto do trabalho desenvolvido pelos pesquisadores e instituições, contribuindo, também para a reforma do sistema de comunicação científica" (Alves, Veiga, 2016).

Com a criação de um canal de comunicação entre as instituições, através do compartilhamento dos arquivos no Google Drive e da lista de discussão, foi possível estabelecer um vínculo de trabalho entre os membros da Rede. Os grupos de trabalho criados permitiram o desenvolvimento de cursos oferecidos aos membros para a qualificação profissional de todos (Figura 2).

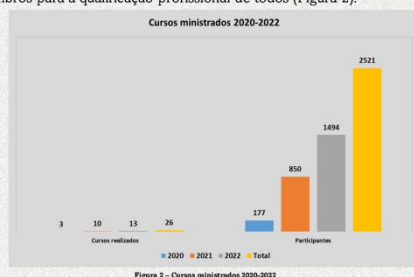


Figura 2 - Cursos ministrados 2020-2022

Também foram realizados eventos com temáticas relevantes sobre Acesso Aberto, Ciência Aberta, Repositórios Digitais, Curadoria e Preservação Digital dentre outros assuntos (Amaro, 2019, 2022). Como resultados, tivemos o I Encontro da Rede Sudeste de Repositórios Institucionais realizado em 2019 com a presença de 700 participantes (Assis, 2019); o II Encontro da Rede Sudeste realizado em 2021 e o I Encontro da Rede Brasileira de Repositórios Digitais, realizado em 2022, ambos no formato remoto e que contou

com aproximadamente 1.000 participantes e mais de 2.000 visualizações.

A maioria dos profissionais que compõem a Rede são formados por Bibliotecários que atuam como gestores dos repositórios, e que atendem a Instituição e seus pesquisadores. Também temos outros profissionais como Analistas de Sistemas, Arquivistas e Técnicos em Bibliotecas.

### Ações realizadas

- ✓ Atualização da listagem dos Contatos das Instituições Participantes da Rede;
- ✓ Realização de reuniões, cursos e eventos;
- ✓ Divulgação das atividades realizadas nas redes sociais;
- ✓ Abertura para participação de instituições das demais regiões nos cursos organizados;
- ✓ Realização de um diagnóstico em 2018 e um novo em 2023 para comparação das informações;
- ✓ Elaboração e revisão das atas das reuniões;
- ✓ Elaboração dos relatórios de atividades como uma forma de "prestação de contas";
- ✓ Ministração de palestras em diferentes instituições como: Marinha do Brasil, Fiocruz, INCA, UFBA; ConfOA, Biredial, etc.;
- ✓ Participação em reuniões realizadas pela Rede Brasileira de Repositórios Digitais, gerenciada pelo IBICT, com a participação das Redes Regionais - Norte, Sul, Centro-Oeste e Nordeste (Assis, 2019);
- ✓ Gerenciamento dos projetos: Certificação de Repositórios Confiáveis (Autoavaliação baseada na norma ISO 16363:2012); Montagem de cursos; Preservação Digital; Política de Indexação para Repositórios; e Auxílio para implementação de Repositórios.

### Considerações finais

Destacamos como estratégicos os principais objetivos da Rede Sudeste, que atualmente conta com 84 instituições e que tem contribuído para o sucesso de suas atividades, como: a criação de repositórios digitais; a busca pelo estabelecimento de uma política de Ciência Aberta para os repositórios de dados de pesquisa; a organização de ações contínuas de aperfeiçoamento dos membros; a ajuda mútua na implantação e gerenciamento dos repositórios; a busca por novas instituições que possuam perfil para integrar a rede; o apoio às ações dos grupos de trabalho e o fortalecimento e implantação das Políticas Institucionais para os Repositórios (Almeida, 2022).

A Rede Sudeste de Repositórios Digitais proporciona um ambiente de colaboração e intercâmbio de conhecimento entre os participantes, visando aproximar as instituições para alcançar os critérios de confiabilidade, credibilidade, autenticidade e qualidade das informações depositadas.

### Bibliografia

- Almeida, A. C. L. de (2022). *Conexões em Acesso Aberto: comunidade de prática em Redes Regionais de Repositórios do Brasil*. 166 f. Dissertação (Mestrado em Ciência da Informação)-Natal, UFRN, 2022. Disponível em: <https://repositorio.ufrn.br/handle/123456789/49878>
- Alves, A. da S., Veiga, V. S. de O. (2016) *Repositórios: conceito, tecnologia e aplicação*. Rio de Janeiro, Fiocruz/Icict. 78 p. Trabalho apresentado no Curso de Acesso à Informação Científica e Tecnológica em Saúde. Modalidade: Qualificação. Disponível em: <http://arca.fiocruz.br/handle/icict/16385>
- Amaro, B. (2019). *A Via Verde do Brasil e a Rede Brasileira de Repositórios Institucionais de Publicações Científicas em Acesso Aberto - RIAA*. In: Barbalho, C. R. S., Inomata, D. O. Galves, J. M., orgs. 1 ed. *A ciência aberta e seus impactos na Região Norte do Brasil*. Manaus: Edua. p. 11-30.
- Amaro, B. (2022). *A Ciência Aberta e a importância da atuação em rede*. Produção: Rede Brasileira de Repositórios Digitais. [S. l.]. 1 arquivo MP4 (45min03s), son., color.
- Assis, T. B. de (2019). *Rede Brasileira de Repositórios e o Impacto dos Trabalhos das Subredes*. In: *Encontro da Rede Sudeste de Repositórios Institucionais, 1., 2019, Rio de Janeiro. Anais... Rio de Janeiro: Fiocruz/Icict/UFRJ, 32 p.*



# Analysis of the Open Data Landscape in Mexico

## **JOANNA ALVARADO URIBE**

INSTITUTE FOR THE FUTURE OF EDUCATION

TECNOLOGICO DE MONTERREY

[joanna.alvarado@tec.mx](mailto:joanna.alvarado@tec.mx)

## **PAOLA MEJÍA ALMADA**

INSTITUTE FOR THE FUTURE OF EDUCATION

TECNOLOGICO DE MONTERREY

[gabriela.almada@tec.mx](mailto:gabriela.almada@tec.mx)

## **ALMA BEATRIZ RIVERA AGUILERA**

BIBLIOTECA FRANCISCO XAVIER CLAVIGERO

UNIVERSIDAD IBEROAMERICANA

[alma.rivera@ibero.mx](mailto:alma.rivera@ibero.mx)

## **BENJAMÍN ALEJANDRO GUERRERO OLVERA**

BIBLIOTECA FRANCISCO XAVIER CLAVIGERO

UNIVERSIDAD IBEROAMERICANA

[benjamin.olvera@ibero.mx](mailto:benjamin.olvera@ibero.mx)

## **MARÍA TERESA VILLALÓN GUZMÁN**

INSTITUTO TECNOLÓGICO DE CELAYA

TECNOLÓGICO NACIONAL DE MÉXICO

[teresa.villalon@itcelaya.edu.mx](mailto:teresa.villalon@itcelaya.edu.mx)

## **MARÍA GUADALUPE VEGA DÍAZ**

BIBLIOTECA DANIEL COSÍO VILLEGAS

EL COLEGIO DE MÉXICO

[guvega@colmex.mx](mailto:guvega@colmex.mx)

## **JOAQUÍN GIMÉNEZ HÉAU**

DIRECCIÓN GENERAL DE REPOSITARIOS UNIVERSITARIOS

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

[joaquin@dgru.unam.mx](mailto:joaquin@dgru.unam.mx)

### **EJE TEMÁTICO**

Datos abiertos

### **ABSTRACT**

The concept of open data originated from the idea of making government data available to anyone. In 2013, the Mexican government launched the National Digital Strategy to encourage the adoption and development of information by implementing five key enablers, including an open data portal to make government data more accessible to citizens. In 2014, the beta version of the [datos.gob.mx](https://datos.gob.mx) portal was launched. Due to the significant potential of data to generate benefits, it has been referred to as the new gold. However, open data also presents some challenges, such as lack of interoperability and reproducibility and the resources needed to provide and maintain open data available to all. Therefore, based on the transversal agenda REMERI-ANUIES-EduTraDi focused on Open Science Ecosystems, the objectives of this research are to identify and analyze institutional repositories promoting Open Data in Mexico in order to document their characteristics and determine possible areas for improvement in such initiatives. This investigation allowed the identification of ten repositories promoting open data in Mexico and the integration of 16 characteristics that helped to describe and analyze the open data landscape, showing that the different open data management efforts carried out by Mexican institutions are mostly consolidated.

### **KEYWORDS**

Open data; repository; Open Data Landscape; Mexico.



# Analysis of the Open Data Landscape in Mexico

XII Conferencia Internacional BIREDIAL-ISTEC, Uruguay, October 18-20, 2023

ALVARADO-URIBE Joanna (Corresponding author)<sup>1</sup>, MEJÍA-ALMADA Paola<sup>1</sup>, RIVERA AGUILERA Alma B.<sup>2</sup>; GUERRERO OLVERA Benjamin A.<sup>2</sup>, VILLALÓN GUZMÁN María Teresa<sup>3</sup>, VEGA DÍAZ María Guadalupe<sup>4</sup>, GIMÉNEZ HÉAU Joaquín<sup>5</sup>

1. Institute for the Future of Education, Tecnológico de Monterrey, Monterrey 64849, Mexico
2. Biblioteca Francisco Xavier Clavigero, Universidad Iberoamericana, Ciudad de México 01219, Mexico
3. Instituto Tecnológico de Celaya, Tecnológico Nacional de México, Mexico
4. Biblioteca Daniel Cosío Villegas, El Colegio de México, Mexico
5. Dirección General de Repositorios Universitarios, Universidad Nacional Autónoma de México, Mexico

joanna.alvarado@tec.mx, gabriela.almada@tec.mx, alma.rivera@ibero.mx, benjamin.olvera@ibero.mx

## Introduction

The concept of open data originated from the idea of making government data available to anyone. In 2013, the Mexican government launched the National Digital Strategy to encourage the adoption and development of information by implementing five key enablers, including an open data portal to make government data more accessible to citizens. However, open data presents some challenges. Therefore, based on the transversal agenda REMERI-ANUIES-EduTraDi focused on Open Science Ecosystems, the objectives of this research are to identify and analyze institutional repositories promoting Open Data in Mexico in order to document their characteristics and determine possible areas for improvement in such initiatives.

Main Theme: Open Data

## Methodology



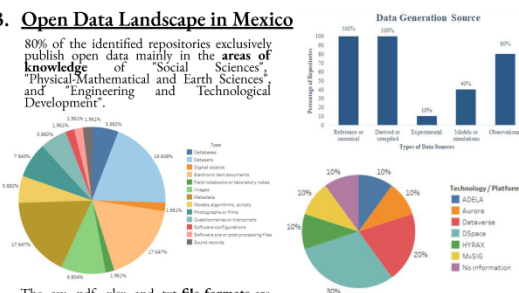
## Results

### 1. Definition of the Collected Characteristics

The 16 characteristics defined for the analysis of the repositories, promoting open data in Mexico, are related to their identification (such as name and Uniform Resource Locator (URL)), the area of knowledge of their content, their typology, numerals, metadata standards, use licenses, among other aspects oriented to data and software.

### 3. Open Data Landscape in Mexico

80% of the identified repositories exclusively publish open data mainly in the areas of knowledge of 'Social Sciences', 'Physical, Mathematical and Earth Sciences', and 'Engineering and Technological Development'.



The .csv, .pdf, .xlsx, and .txt file formats are published in open data repositories.

On the one hand, the Dublin Core Metadata Initiative (DCMI) is the most followed by repositories by incorporating its Dublin Core (DC) vocabulary. On the other hand, the Creative Commons (CC) licenses were the most widely used in repositories.

## Conclusions

- This investigation allowed the identification of ten repositories promoting open data in Mexico and the integration of 16 characteristics that helped to describe and analyze the open data landscape (until May 12, 2023).
- The results of this research show that the different open data management efforts carried out by Mexican institutions are mostly consolidated since they incorporate metadata standards, platforms, and use licenses that enable free data access, use, and sharing.
- Some areas for improvement were identified:
  - Establish an agreement on the information to be provided and its structuring for a better description, analysis, and evaluation of the data and the repositories.
  - Promote the integration of open data in a unique repository or in different repositories where it is sought to unify metadata standards and the hosting technology or platform.
  - Strengthen the adoption of laws and agreements on the protection of personal data and the treatment of sensitive data.
  - Encourage the integration of more private institutions in Mexico's Open Data initiative.

## References

Open Knowledge Foundation. (2015, November). Open Definition. Retrieved May 13, 2023, from DEFINING OPEN IN OPEN DATA, OPEN CONTENT AND OPEN KNOWLEDGE: <https://opendefinition.org/>

CONACYT. (2022). TÉRMINOS DE REFERENCIA. Retrieved May 13, 2023, from Convocatoria de Ciencia Básica y/o Ciencia de Frontera: [https://conacyt.mx/wp-content/uploads/convocatoria/ciencia\\_basica\\_frontera/paradigma\\_y\\_cierre/terminos\\_referencia\\_ab\\_ciencia\\_2022\\_V1.pdf](https://conacyt.mx/wp-content/uploads/convocatoria/ciencia_basica_frontera/paradigma_y_cierre/terminos_referencia_ab_ciencia_2022_V1.pdf)

Digital Government Unit, M. C. (2015, September 15). Digital government toolkit. Retrieved May 14, 2023, from Mexico: National Digital Strategy: <https://www.oecd.org/gov/mexico-digital-strategy.pdf>

Gohm, A., & Sandhu, R. (2020, January). Open data repositories. Retrieved May 13, 2023, from Current risks and opportunities: <https://arxiv.org/abs/2001.01001v1>

LEARN project. L. A. (2020, December 18). Gestión de datos de investigación. Retrieved May 13, 2023, from Gestión de datos de investigación: <https://biblioteca-cpaq.ing.unam.mx/gestion-de-datos-de-investigacion>

Schnee, D., Godey, G., Hahnel, M., Zhou, Y., Jiang, L., Chandrasekharan, L., ... & Dey, L. (2022, October 18). The State of Open Data 2022. doi:<https://doi.org/10.6084/m9.figshare.2127884.v5>

### 2. Repositories Promoting Open Data in Mexico

The identification of Mexican repositories promoting open data was carried out from September 9, 2022, to November 2, 2022.

No.	Repository Name	Repository URL	Institution	Sector	Open Data Exclusive
1	Portal de Datos Abiertos UNAM	<a href="https://datosabiertos.unam.mx/">https://datosabiertos.unam.mx/</a>	Universidad Nacional Autónoma de México (UNAM) / Dirección General de Repositorios Universitarios	Public	Yes
2	Data Hub del Tecnológico de Monterrey	<a href="https://datahub.tec.mx/">https://datahub.tec.mx/</a>	Tecnológico de Monterrey	Private	Yes
3	Datos Abiertos de México	<a href="https://datos.gob.mx/">https://datos.gob.mx/</a>	Gobierno de México	Public	Yes
4	Laboratorio Nacional de Políticas Públicas	<a href="http://datos.cide.edu/community-list">http://datos.cide.edu/community-list</a>	Laboratorio Nacional de Políticas Públicas	Public	Yes
5	CIMMYT Research Data & Software Repository Network	<a href="https://data.cimmyt.org/datasets/root?q=sort:dataset&amp;order=asc&amp;page=1">https://data.cimmyt.org/datasets/root?q=sort:dataset&amp;order=asc&amp;page=1</a>	International Maize and Wheat Improvement Center	Public and private	Yes
6	Repositorio Digital del Servicio Sismológico Nacional	<a href="http://repositorio.ssn.unam.mx/sgn/">http://repositorio.ssn.unam.mx/sgn/</a>	UNAM / Servicio Sismológico Nacional / Instituto de Geología	Public	Yes
7	Datos abiertos - INEGI	<a href="https://www.inegi.org.mx/datosabiertos/">https://www.inegi.org.mx/datosabiertos/</a>	Instituto Nacional de Estadística y Geografía (INEGI)	Public	Yes
8	Repositorio de Documentación sobre Desapariciones en México (RDDM)	<a href="https://rddm.mx/">https://rddm.mx/</a>	Center for Research Libraries; El Colegio de México; Universidad Iberoamericana; UNAM / Instituto de Investigaciones Jurídicas	Public	Partially
9	Datos Abiertos - SSA/DGIS	<a href="http://www.dgis.salud.gob.mx/contenidos/basesdedatos/DatosAbiertos_gobmx.html">http://www.dgis.salud.gob.mx/contenidos/basesdedatos/DatosAbiertos_gobmx.html</a>	Secretaría de Salud (SSA) / Dirección General de Información en Salud (DGIS)	Public	Yes
10	Repositorio IBERO	<a href="https://ri.ibero.mx/">https://ri.ibero.mx/</a>	Universidad Iberoamericana	Private	Partially



**Acknowledgments.** Authors thank Gerardo Castañeda Garza for his collaboration in the early phases of the project, as well as Rosalva Vázquez Tapia, Antonio Razo and Martín Adalberto Jena Espinoza de los Monteros for the coordination and support within the framework of the transversal agenda REMERI-ANUIES-EduTraDi to which this research belongs.

# Análise da viabilidade de associação dos planos de gestão de dados acionáveis por máquina com o ciclo de vida dos dados no ecossistema de pesquisa brasileira

## LUCIA HELENA CUNHA VIDAL

UFRGS

[luciahcvidal@hotmail.com](mailto:luciahcvidal@hotmail.com)

## RENE FAUSTINO GABRIEL JUNIOR

UFRGS

[rene.gabriel@ufrgs.br](mailto:rene.gabriel@ufrgs.br)

## CATERINA MARTA GROPOSO PAVÃO

UFRGS

[caterina@cpd.ufrgs.br](mailto:caterina@cpd.ufrgs.br)

### RESUMO

Discute a importância dos dados de pesquisa e destaca a necessidade de plano de gestão de dados adequado para garantir disponibilidade, acesso, processamento e reuso dos dados, no eixo temático 2: dados abertos. Expõe que o problema dos planos de gestão de dados não estão integrados com outros atores no ecossistema de pesquisa, dificultando o acompanhamento do ciclo de vida dos dados. Propõe analisar a viabilidade da utilização de ferramentas interoperáveis com diversas bases de dados para tornar os planos de gestão de dados acionáveis por máquina e integrados com o ciclo de vida dos dados no ecossistema BrCris. É esperado contribuir à prospecção de planos de gestão de dados acionáveis por máquina aprimorando sua implantação e uso no Brasil. De natureza básica, abordagem qualitativa e procedimentos de pesquisa bibliográfica e documental, será descritiva e explicativa. A coleta de dados será realizada em bases de dados bibliográficas nacionais e internacionais da Ciência da Informação, Ciência da Computação e Administração. Apresenta ferramentas para elaboração de planos de gestão de dados e o padrão comum acionável por máquina da Research Data Alliance, que contempla as etapas do ciclo de vida dos dados e os atores de um ecossistema de pesquisa. Analisa desafios e oportunidades para implementação dos

planos de gestão de dados acionáveis por máquina no Brasil. É esperado que esta pesquisa contribua para a identificação e prospecção de modelos de maDMPs para pesquisadores, compreendendo suas características, estruturas e abordagens para gestão de dados de pesquisa, além fornecer uma visão abrangente dos modelos existentes e suas contribuições para integração com ecossistemas de pesquisa.

## XII Conferência Internacional BIREDIAL-ISTEC

Uruguai, 18-20 de outubro de 2023

### título

Análise da viabilidade de associação dos planos de gestão de dados acionáveis por máquina com o ciclo de vida dos dados no ecossistema de pesquisa brasileira

### problema

Qual a viabilidade de associação dos planos de gestão de dados acionáveis por máquina com o ciclo de vida dos dados no ecossistema da pesquisa brasileira?

### objetivo

Analisar conceitualmente a viabilidade de associação do plano de gestão de dados acionáveis por máquina com o ciclo de vida dos dados no ecossistema da pesquisa brasileira.

### metodologia

De natureza básica, abordagem qualitativa e procedimentos de pesquisa bibliográfica e documental, será descritiva e explicativa.



### resultados parciais e esperados

Espera-se que esta pesquisa contribua para a identificação e prospecção de modelos de *maDMPs* para pesquisadores, compreendendo suas características, estruturas e abordagens para gestão de dados de pesquisa, além fornecer uma visão abrangente dos modelos existentes e suas contribuições para integração com ecossistemas de pesquisa. Ambiciona-se promover a conscientização e o envolvimento dos atores do ecossistema, fortalecendo a colaboração e troca de conhecimento no contexto brasileiro. Um dos principais resultados esperados é a verificação da aplicabilidade dos *maDMPs* no contexto BrCris, fornecendo assim, informações relevantes sobre a viabilidade para implementação bem-sucedida dos *maDMPs* no contexto brasileiro.

### autores

**Lucia Helena Cunha Vidal**, [luciahcvidal@hotmail.com](mailto:luciahcvidal@hotmail.com), UFRGS  
**Rene Faustino Gabriel Junior**, [rene.gabrieleufrgs.br](mailto:rene.gabrieleufrgs.br), UFRGS  
**Caterina Marta Groposo Pavão**, [caterina@cpd.ufrgs.br](mailto:caterina@cpd.ufrgs.br), UFRGS



This is an open-access article distributed under the terms of the Creative Commons Attribution License.



# Infraestructura para soportar la Ciencia Abierta en la Universidad del Rosario

## HUMBERTO BLANCO CASTILLO

UNIVERSIDAD DEL ROSARIO (COLOMBIA)

[humberto.blanco@urosario.edu.co](mailto:humberto.blanco@urosario.edu.co)

## MALGORZATA LISOWSKA

UNIVERSIDAD DEL ROSARIO (COLOMBIA)

[margarita.lisowska@urosario.edu.co](mailto:margarita.lisowska@urosario.edu.co)

## RESUMEN

La Universidad del Rosario en Colombia ha implementado un modelo de Ciencia Abierta que reconoce la importancia de contar con una infraestructura abierta e interoperable que soporte de manera transversal todo el modelo. Para lograr esto, ha integrado las plataformas: repositorio Institucional EdocUR, repositorio de datos de investigación UR, portal de revistas UR y el portal de fortalezas institucionales (HUB-UR:Services and Experts Finder), las cuales son capaces de interoperar entre sí e integrarse con fuentes internas y externas. Para garantizar la interoperabilidad, fue necesario realizar la normalización de metadatos en las fuentes internas y garantizar el cumplimiento de estándares. Los resultados de esta iniciativa incluyen: la inserción en indexadores y redes académicas a nivel nacional e internacional; el fortalecimiento de la visibilidad académica en la web; la normalización de la presencia digital mediante el uso de estándares como ORCID y DOI; la normalización de los datos en los sistemas fuentes de la universidad; la creación de servicios de valor agregado en el HUB-UR: Services & Experts Finder.

## PALABRAS CLAVES

Interoperabilidad; infraestructura abierta; Ciencia Abierta. Interoperability; Open Access infrastructure; Open Science.



## BIREDIAL-ISTEC

XII CONFERENCIA INTERNACIONAL BIREDIAL-ISTEC 2023

INFRAESTRUCTURA PARA SOPORTAR LA CIENCIA ABIERTA EN LA UNIVERSIDAD DEL ROSARIO (COLOMBIA)

Malgorzata Lisowska (margarita.lisowska@urosario.edu.co)

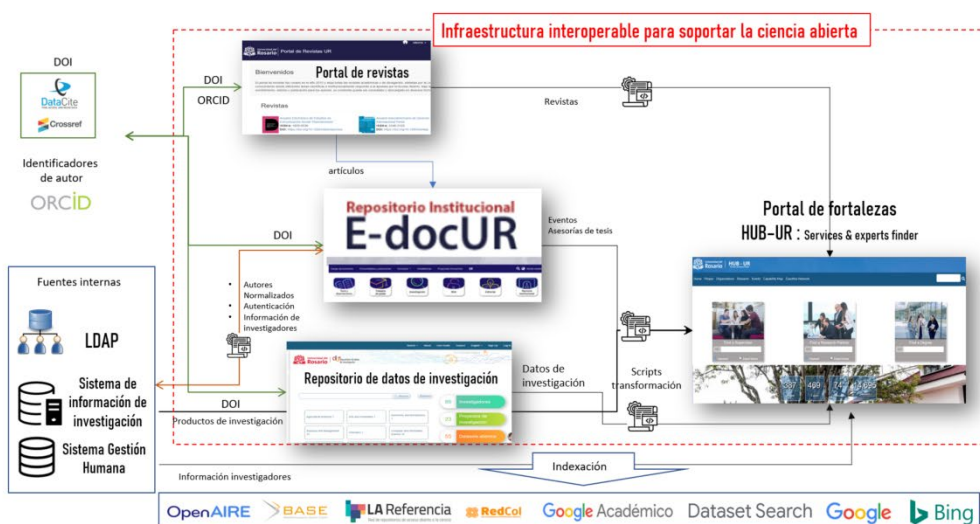
Humberto Blanco (humberto.blanco@urosario.edu.co)

Universidad del Rosario - Colombia

El modelo de Ciencia Abierta de la Universidad del Rosario reconoce la importancia de contar con una infraestructura tecnológica interoperable, para la materialización y el sostenimiento del mismo. Es por eso que un pilar transversal de este modelo está dedicado a recursos y plataformas, que conforman el ecosistema institucional.



Modelo de ciencia abierta UR. Tomado de: <https://cienciaabierta.urosario.edu.co/nuestro-modelo.php>, Universidad del Rosario. (2020)



### RESULTADOS

La creación de la infraestructura tecnológica 100% interoperable, para soportar a la Ciencia Abierta en la Universidad del Rosario, ha permitido:

- La inserción en indexadores, redes académicas nacionales e internacionales.
- Fortalecimiento de la visibilidad académica en la web de la Universidad y sus miembros.
- Normalización de la presencia digital de los investigadores y su producción, gracias al uso de los estándares internacionales como ORCID, DOI, etc.
- Normalización de los datos en los sistemas fuente de la Universidad.
- Creación de los servicios de valor agregado como: busca un tutor de tesis, encuentra un partner de investigación o busca un programa académico en el HUB-UR: Services & Experts Finder .

Este obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional.



# Periódicos científicos: normativas para abertura de dados

## LARISSA B. B. DRUMOND

UNIVERSIDADE FEDERAL DE GOIÁS

[larissa.barbara@ufg.br](mailto:larissa.barbara@ufg.br)

## LAURA V. R. REZENDE

UNIVERSIDADE FEDERAL DE GOIÁS

[laura\\_rezende@ufg.br](mailto:laura_rezende@ufg.br)

### RESUMO

Este trabalho é um dos desdobramentos de uma pesquisa de mestrado realizada na Universidade Federal de Goiás, Brasil. O recorte selecionado refere-se à abertura dos dados científicos adjacentes aos artigos publicados pelos periódicos, com o viés da Ciência Aberta que, segundo a Recomendação de 2022 da UNESCO prevê a abertura do conhecimento científico, tornando-o acessível e reutilizável para todos. Compreende-se que uma abertura efetiva requer o estabelecimento de normativas que culminem no compartilhamento de dados que sejam localizáveis, acessíveis, interoperáveis com outros sistemas e reutilizáveis, atendendo aos princípios FAIR. Assim, realizaram-se análises da certificação Core Trust Seal, que apresenta requisitos de repositórios sustentáveis e confiáveis, e das Diretrizes Transparency and Openness Promotion (TOP), que objetiva transparência e abertura nas políticas e práticas editoriais de periódicos. Conclui-se que as ações para se requerer dos autores o compartilhamento dos dados perpassam por: 1) definições preliminares, como decisões regulatórias institucionais e estabelecimento de normativas internas dos periódicos; 2) ações de funcionamento, como as instruções aos autores sobre preparo, envio ou depósito dos dados, curadoria e acompanhamento das estatísticas de acesso e uso dos dados, independente do repositório de dados escolhido para armazenamento e publicação dos dados.

### PALAVRAS CHAVES

Ciência Aberta; comunicação científica; dados abertos. Open Science; scientific communication; open data.

CONFERENCIA INTERNACIONAL  
**BIREDIAL-ISTEC** '23  
 18-20 de octubre de 2023 MONTEVIDEO • URUGUAY

UFMG  
 UNIVERSIDADE FEDERAL DE GOIÁS

# PERIÓDICOS CIENTÍFICOS: normativas para abertura de dados

DRUMOND, Larissa B. B - larissa.barbara@ufg.br  
 REZENDE, Laura V. R. - laura\_rezende@ufg.br  
 Universidade Federal de Goiás, Brasil

- Resumo** O estudo refere-se à abertura dos dados científicos adjacentes aos manuscritos publicados pelos periódicos, com o viés da Ciência Aberta.
- Objetivo** Estabelecer as ações e normativas que culminem no compartilhamento de dados científicos que sejam localizáveis, acessíveis, interoperáveis com outros sistemas e reutilizáveis, atendendo aos princípios FAIR.
- Método** Análise da certificação *Core Trust Seal* e das *Diretrizes Transparency and Openness Promotion (TOP)*.

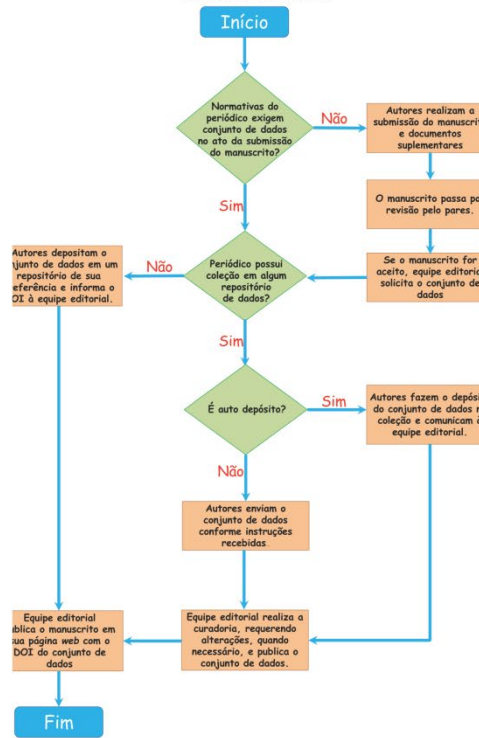
**Resultados:**

Figura 1 - Ações para implantação de abertura de dados de periódicos científicos

	AÇÕES/ DEFINIÇÕES PRELIMINARES DO PERIÓDICO	AÇÕES DE FUNCIONAMENTO DO PERIÓDICO
<b>Estrutura Organizacional e Regulamentar</b>	<ol style="list-style-type: none"> <li>Caso haja um repositório institucional de dados, definições sobre a criação da coleção, licenças aplicáveis e tipos de dados aceitos são de responsabilidade da instituição de origem do periódico.</li> <li>Definição da missão da coleção de dados abertos do periódico;</li> <li>Elaboração de um modelo de Declaração de Disponibilidade de Dados e instruções para preparação dos dados.</li> </ol>	<ol style="list-style-type: none"> <li>Acompanhamento do acesso e uso da coleção (estatísticas).</li> </ol>
<b>Estrutura Técnica e Tecnológica</b>	<ol style="list-style-type: none"> <li>Criação e configuração da coleção no repositório de dados (institucional ou externo) conforme recursos disponíveis;</li> <li>Definição dos membros da equipe editorial responsáveis pela coleção de dados;</li> <li>Elaboração de um formulário padrão (template) de metadados a ser preenchido pelos autores;</li> <li>Criação de um fluxo de trabalho para a preparação e envio dos dados por parte dos autores;</li> <li>Criação de um fluxo de trabalho para o depósito e publicação.</li> </ol>	<ol style="list-style-type: none"> <li>Recebimento dos conjuntos de dados e respectivos metadados enviados pelos autores (no caso do depósito ser realizado pela equipe editorial)</li> <li>Conferência e adequações do conjunto de dados (curadoria);</li> <li>Publicação do conjunto de dados.</li> </ol>
<b>Estrutura Formativa e de Divulgação</b>	<ol style="list-style-type: none"> <li>Elaboração de um plano de divulgação da coleção de dados do periódico.</li> <li>Documento orientando autores sobre descrição do contexto do projeto de pesquisa que originou o artigo submetido na periódica, anonimização e de-identificação dos dados, dentre outros fatores a serem considerados antes de divulgar os dados.</li> </ol>	<ol style="list-style-type: none"> <li>Orientações aos autores quanto ao depósito dos dados;</li> <li>Elaboração de ações de divulgação da coleção e da política de fortalecimento da Ciência Aberta (no site do periódico, página da coleção de dados e redes sociais).</li> </ol>

Fonte: elaborado pelas autoras.

Figura 2 - Fluxograma para compartilhamento de dados científicos



Fonte: elaborado pelas autoras.

Referências



# El impacto de los Identificadores Persistentes (PIDs) en el Repositorio de Datos Académicos RDA-UNR

## **ANALÍA SALAZAR**

UNIVERSIDAD NACIONAL DE ROSARIO

[analia.salazar@unr.edu.ar](mailto:analia.salazar@unr.edu.ar)

## **PAULINA FREÁN**

UNIVERSIDAD NACIONAL DE ROSARIO

[paulina.frean@unr.edu.ar](mailto:paulina.frean@unr.edu.ar)

## **GISELA CHIAPPERO**

UNIVERSIDAD NACIONAL DE ROSARIO

[gisela.chiappero@unr.edu.ar](mailto:gisela.chiappero@unr.edu.ar)

## **DOLORES QUINTANA**

UNIVERSIDAD NACIONAL DE ROSARIO

[doloresquintanaunr@gmail.com](mailto:doloresquintanaunr@gmail.com)

## **PAOLA C. BONGIOVANI**

UNIVERSIDAD NACIONAL DE ROSARIO

[pbongiovani@gmail.com](mailto:pbongiovani@gmail.com)

## **AGUSTÍN ALFIERI**

UNIVERSIDAD NACIONAL DE ROSARIO

[agustin.alfieri@unr.edu.ar](mailto:agustin.alfieri@unr.edu.ar)

## **RESUMEN**

Se presenta la implementación de Identificadores Persistentes (PIDs) en el Repositorio de Datos Académicos RDA-UNR (<https://dataverse.unr.edu.ar>). Los problemas detectados fueron la necesidad de identificación única de los conjuntos de datos, de los archivos, de los investigadores y de la Universidad para asegurar la correcta atribución de autoría y facilitar la referenciación precisa de los elementos, además de asegurar la integridad y recuperación de los datos, su correcta citación y vinculación con otros recursos académicos. Se examinan específicamente los identificadores ORCID, DOI y ROR, y se evalúa su presencia y utilización en los metadatos de los conjuntos de datos.

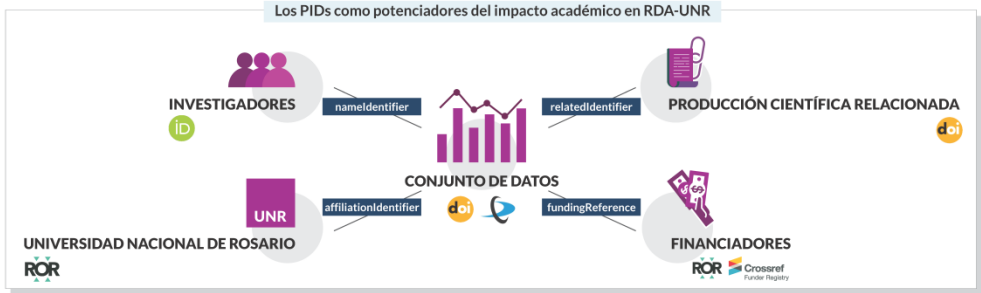
## **PALABRAS CLAVE**

Repositorio de datos; conjuntos de datos; metadatos; identificadores persistentes. Data Repository; dataset; metadata; persistent identifiers.

# El impacto de los Identificadores Persistentes (PIDs) en el Repositorio de Datos Académicos RDA-UNR

<https://dataverse.unr.edu.ar>

Eje temático | 2. Datos Abiertos



**Autores**  
 Salazar, A.<sup>1</sup>, Freán, P.<sup>1,2</sup>, Chiappero, G.<sup>1</sup>, Quintana, D.<sup>1</sup>, Bongiovani, P. C.<sup>1,2</sup>, Alfieri, A.<sup>1</sup>  
 1. Universidad Nacional de Rosario. Unidad de Gestión de Acceso Abierto.  
 2. Universidad Nacional de Rosario. Facultad de Humanidades y Artes.  
 Centro de Estudios Interdisciplinarios en Bibliotecología y Ciencias de la Información.

**Palabras clave**  
 Repositorio de datos, Conjuntos de datos, Metadatos, Identificadores Persistentes, Data Repository, Dataset, Metadata, Persistent Identifiers.

**Introducción**  
 Se presenta la implementación de Identificadores Persistentes (PIDs) en el Repositorio de Datos Académicos RDA-UNR (<https://dataverse.unr.edu.ar>). Los problemas detectados fueron la necesidad de identificación única de los conjuntos de datos, de los archivos, de los investigadores y de la Universidad para asegurar la correcta atribución de autoría y facilitar la referenciación precisa de los elementos, además de asegurar la integridad y recuperación de los datos, su correcta citación y vinculación con otros recursos académicos. Se examinan específicamente los identificadores ORCID, DOI y ROR, y se evalúa su presencia y utilización en los metadatos de los conjuntos de datos.

**Metodología**  
 La metodología utilizada consiste en la implementación de PIDs en los datos académicos almacenados en el repositorio RDA-UNR, la contratación de DataCite para la provisión de DOIs para el repositorio de datos por medio del Consorcio Dataverse y la contratación de Crossref para la adquisición de DOIs para las revistas del Portal de Revistas UNR, la formación de investigadores en cuanto al valor de los PIDs y, por último, la curación de datos por expertos. Se examinaron los metadatos de los conjuntos de datos en busca de la presencia de los indicadores persistentes ORCID, DOI y ROR. Se solicitó a cada uno de los autores de los conjuntos de datos del RDA-UNR que no contaban con identificador ORCID, su registro haciendo énfasis en la importancia, tanto en las consultas virtuales de curación de datos como en webinars específicos, presentaciones en eventos y jornadas, capacitaciones presenciales en las distintas Unidades Académicas de la Universidad y mediante difusión por redes sociales, mails mensuales y sitios web institucionales.

**Resultados**  
 La implementación de PIDs en el Repositorio de Datos Académicos RDA-UNR ha mejorado significativamente la identificación y citación persistentes de los conjuntos de datos almacenados y, por ende, su impacto. Los resultados muestran que todos los conjuntos de datos y archivos del repositorio cuentan con identificadores DOIs, lo que permite su identificación única y permanente y, a su vez, desde DataCite se recuperan las citas obtenidas en Crossref. El 72% de autores de los conjuntos de datos publicados cuenta con identificadores ORCID en sus metadatos, lo que facilita la atribución adecuada de los datos. Ningún dataset cuenta con el identificador ROR, debido a que Dataverse se encuentra desarrollando su implementación, lo que mejoraría la afiliación institucional y la interoperabilidad con otros recursos. Para difundir el identificador ROR de la Universidad Nacional de Rosario se utiliza el sitio web RDA-UNR Info (<https://dataverse-info.unr.edu.ar>) y otros sitios como DataCite Commons (<https://commons.datacite.org/ror.org/02tphfq59>).

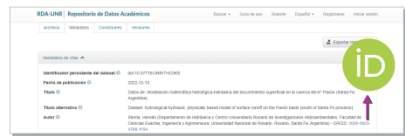
**Conclusión**  
 En conclusión, el póster destaca que los PIDs ayudan a abordar los desafíos de identificación, atribución, enlace y gestión de datos en la Universidad Nacional de Rosario, mejorando la accesibilidad, la visibilidad, la interoperabilidad de los recursos y la colaboración en contribuciones científicas.



<https://doi.org/10.57715/UNR/THCOKS>



<https://orcid.org/0000-0003-4788-1194>



<https://doi.org/10.57715/UNR/THCOKS>



<http://commons.datacite.org/ror.org/02tphfq59>



Comité de Acceso Abierto UNR  
 Unidad de Gestión de Acceso Abierto UNR  
 Facultad de Humanidades y Artes. Centro de Estudios Interdisciplinarios en Bibliotecología y Ciencias de la Información

Email: [repositoriodatos@unr.edu.ar](mailto:repositoriodatos@unr.edu.ar)

Más info



# Práticas de compartilhamento de dados pelos pesquisadores PQ da área de Física e Astronomia do CNPQ

**RENÉ FAUSTINO GABRIEL JUNIOR**

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

[rene.gabriel@ufrgs.br](mailto:rene.gabriel@ufrgs.br)

**TATYANE GUEDES MARTINS DA SILVA**

**THIAGO MAGELA RODRIGUES DÍAS**

## RESUMO

Compartilhar os dados de pesquisa significa promover a transparência, promover a colaboração científica, evitar a duplicação desnecessária de esforços e possibilitar a reprodutibilidade das pesquisas. Faz-se a importância de estudar as práticas de compartilhamento de dados de pesquisa pelas diversas áreas do conhecimento de forma a ter um entendimento de suas práticas e a cultura do compartilhamento de dados. O uso dos identificadores persistentes nas publicações científicas, repositórios científicos e nos conjuntos de dados está possibilitando o cruzamento de informações nas diversas bases de dados. O uso de robôs de coleta também tem facilitado o manuseio desses identificadores em diversas bases de dados.



# Práticas de compartilhamento de dados pelos pesquisadores PQ da área de Física e Astronomia do CNPQ

Rene Faustino Gabriel Junior (1), Tatyane Guedes Martins da Silva (2), Thiago Magela Rodrigues Dias (3)

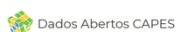
## Introdução

Compartilhar os dados de pesquisa significa promover a transparência, promover a colaboração científica, evitar a duplicação desnecessária de esforços e possibilitar a reprodutibilidade das pesquisas. Faz-se a importância de estudar as práticas de compartilhamento de dados de pesquisa pelas diversas áreas do conhecimento de forma a ter um entendimento de suas práticas e a cultura do compartilhamento de dados.

O uso dos identificadores persistentes nas publicações científicas, repositórios científicos e nos conjuntos de dados está possibilitando o cruzamento de informações nas diversas bases de dados. O uso de robôs de coleta também tem facilitado o manuseio desses identificadores em diversas bases de dados.

## Objetivo

Identificar as práticas de compartilhamento de dados por pesquisadores PQ da área de física e astronomia do CNPQ



## Metodologia

Para identificar as práticas de compartilhamento de dados de pesquisa por pesquisadores das áreas de física e astronomia parte-se do arquivo BR-CAPES-COLSUCUP-DOCENTE-2020-2021-11-10 disponibilizado no Portal Dados Abertos CAPES. Extraí os pesquisadores da área de avaliação 3 (física e astronomia) que tiveram ou tem bolsa produtividade PQ do CNPQ resultando em uma lista de 1.078 pesquisadores.

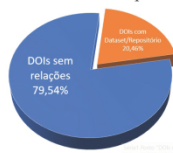
Para possibilitar a análise foram coletadas das produções científicas, foi necessário a utilização de uma API para coleta de dos arquivos XML no LattesExtrator de cada pesquisador. Esta coleta resultou na identificação de 107.444 artigos, sendo 19.377 (18,0%) desses trabalhos sem DOI. Com este corpus de 88.607 (82,0%) trabalhos foi construído um script em Python, para consulta via API na base do ScholeXplorer. Para não sobrecarregar o servidor, as consultas foram realizadas com intervalos de 1 seg. cada, resultando aproximadamente 24 horas para finalizar as consultas. Todos os dados foram armazenando em uma base de dados MySQL.

Com os dados pode-se verificar a relação entre os DOIs, o que o ScholeXplorer chama de "Source" e "Target" com seu "Relationship". Os resultados dessas relações foram inseridas em um dataframe para serem analisados.

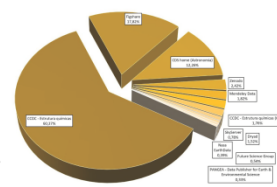
## Resultados

Com o processamento dos dados foi possível verificar que os DOIs "target" coletados no ScholeXplorer podem ser divididos em dois grandes grupos, um referente aos conjuntos de dados de pesquisa e um segundo relacionado a publicações pré-print e suporte a publicação científica tradicional das revistas.

Também foi possível observar que os pesquisadores da física utilizam o arXiv. Sendo um repositório de arquivo para preprints eletrônicos de artigos científicos nos campos da matemática, física, ciência da computação, biologia quantitativa e estatística que podem ser acessados via internet. Em muitos campos da matemática e da física, quase todos os artigos estão no arXiv. No estudo, observou-se que 88,0% dos dados em preprint recuperados estavam relacionados a este repositório, ou seja, muito utilizados pelos pesquisadores da física.



Observa-se que na Figura 1, 20,40% dos DOIs dos artigos tem relacionamento com algum dataset ou repositório. Os dados foram gerados



Com base nos DOIs dos compartilhamentos permitiram identificar as fontes dos dados de pesquisas compartilhados, observou-se que a área de física e astronomia tiveram o maior depósito de dados no Cambridge Crystallographic Data Centre (CCDC) especializada em dados de química estrutural e desenvolvimento de materiais e ciências da vida para uso em descoberta farmacêutica.

O segundo repositórios de dados mais frequente é o Figshare, repositório científico online, aberto e gratuito que permite aos pesquisadores acadêmicos preservar e compartilhar todos os dados e resultados gerados em pesquisa, de documentos e procedimentos operacionais, padrões e protocolos. O Figshare foi criado em 2011 e é um dos vários negócios de portfólio apoiados pela Digital Science3, uma subsidiária da Springer Nature.

Com 12,26% dos depósitos de dados, o Centre de Données astronomiques de Strasbourg (CDS home), é um centro de dados que recolhe e distribui informações astronômicas. Foi fundado em 1972 com o nome de Centre de Données Stellaires. Os serviços on-line atualmente fornecidos pelo CDS são um banco de dados de objetos astronômicos, serviço para catálogos astronômicos, e atlas celeste interativo e banco de dados de imagens. Também foram identificadas depósitos nos repositórios de dados do Zenodo (2,4%), Mendley Data (1,82%), Dryad (1,52%).

## Considerações finais

Este estudo foi utilizado para avaliar a viabilidade e confiança nas informações do cruzamento de dados entre a publicações científica e os conjuntos de dados. Entretanto, observou-se que a prática de registrar DOIs em repositórios de preprint é recorrente pelos pesquisa da área da Física. Considera-se que o objetivo do estudo foi desenvolvido, entretanto necessita de refinamento na metodologia para melhor identificação das variáveis e as especificidades de cada área do conhecimento.





# Interoperabilidad automática gracias al uso de PIDs - ORCID

**ANA CARDOSO**

ORCID

## RESUMEN

El objetivo de ORCID es conectar a los investigadores con sus resultados de investigación, y dado que los datos de ORCID son accionables por máquina, el principio FAIR de interoperabilidad es donde ORCID realmente brilla. La infraestructura interoperable de ORCID puede ayudar a acelerar el descubrimiento de conocimientos y aumentar la integridad, la transparencia y la reproducibilidad de la investigación al fomentar los principios de datos FAIR y las prácticas de Ciencia Abierta a través de identificadores persistentes y datos estandarizados y de acceso abierto.

# id Interoperabilidad Automática gracias al uso de PIDs - ORCID

XII Conferencia Internacional BIREDIAL-ISTEC, Uruguay, 18-20 octubre 2023

Ana Cardoso ORCID - Senior Engagement Lead <https://orcid.org/0000-0002-8631-3838>

## ORCID juega un papel central en el ecosistema de investigación al conectar los sistemas de información académicos



Los ID de ORCID son abiertos e interoperables, pueden incluirse en cualquier sistema o plataforma, y son persistentes a través de las disciplinas, fronteras y a lo largo del tiempo

**Identificadores para Organizaciones**

Los identificadores organizacionales son necesarios para que los investigadores se conecten de manera segura a sus organizaciones de afiliación y viceversa, incluso si las organizaciones cambian de nombre o lugar

**Identificadores para Trabajos/ Contribuciones**

Los PIDs para las contribuciones son necesarios para dar visibilidad y para su interoperabilidad, ya que permiten conexiones persistentes entre diferentes sistemas y plataformas. ORCID soporta el uso de identificadores diferentes para contribuciones, como son los DOI

El ciclo virtuoso de interoperabilidad representa cómo las infraestructuras interoperan entre sí apoyándose en identificadores persistentes (PID)

**ORCID** es una organización sin fines de lucro que ayuda a crear un mundo en el que todos los que participan en la investigación y la innovación estén identificados de forma única y conectados a sus contribuciones y afiliaciones, atravesando disciplinas, fronteras y tiempo. En octubre de 2012, ORCID lanzó su registro abierto de identificadores de personas para investigadores y académicos. A junio 2023, más de 9.5 millones de personas han registrado y llenado su registro ORCID. Más de 1300 organizaciones de todo el mundo se han unido a ORCID como miembros institucionales para incorporar IDs en sus sistemas y flujos de trabajo, y editoriales y organismos de financiación están comenzando a requerir e implementar IDs en sus procesos

# Taxonomía de la ciencia abierta: revisada y ampliada

## **LÚCIA DA SILVEIRA**

UNIVERSIDADE FEDERAL DE SANTA CATARINA (UFSC)

## **REMEDIOS MELERO**

INSTITUTO DE AGROQUÍMICA Y TECNOLOGÍA DE ALIMENTOS

## **NIVALDO CALIXTO RIBEIRO**

UNIVERSIDADE FEDERAL DE MINAS GERAIS

## **ANDREA MORA-CAMPOS**

UNIVERSIDAD NACIONAL

## **PRISCILA MACHADO BORGES SENA**

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

## **FABIANO COUTO CORRÊA DA SILVA**

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

## **JULIANA FACHIN**

GRUPO DE INVESTIGAÇÃO CIÊNCIA ABERTA LATAM

## **DANIEL FERNANDO PIRAQUIVE-PIRAQUIVE**

UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS

## **JORGE POLANCO CORTÉS**

UNIVERSIDAD DE COSTA RICA

## **RONALDO FERREIRA ARAÚJO**

UNIVERSIDADE FEDERAL DE ALAGOAS

## **ALEJANDRO URIBE TIRADO**

UNIVERSIDAD DE ANTIOQUIA

## **JULIO SANTILLÁN-ALDANA**

UNIVERSIDAD TECNOLÓGICA DEL PERÚ

## **ANDRÉS MAURICIO ENCISO BETANCOURT**

UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS

### **RESUMEN**

**Objetivo:** revisar las terminologías y aplicaciones de la taxonomía de la Ciencia Abierta, para construir una versión más completa, que represente el conocimiento en torno al tema, de acuerdo con el escenario actual de la comunicación científica y las recomendaciones de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (Unesco). **Método:** se trata de una investigación exploratoria con un enfoque deductivo. El primer paso fue la revisión de las taxonomías, con las 12 personas dedicadas a la investigación, que se reunieron semanalmente, para las discusiones conceptuales y epistemológicas relacionadas con la Ciencia Abierta y las definiciones metodológicas y de procedimiento para la realización del estudio. **Resultados:** como resultado de los análisis, se elaboró una taxonomía para ser evaluada por las personas expertas. Para ello, se envió a 68 personas expertas un cuestionario con preguntas abiertas sobre cada eje principal de la taxonomía. Se obtuvieron 21 respuestas que cooperaron con la modelización y exposición de los términos de la nueva propuesta de taxonomía. La nueva taxonomía tiene 10 componentes de nivel principal y un total de 96 etiquetas. **Conclusiones:** la percepción de los expertos puso de manifiesto un panorama congruente con las recomendaciones de la Unesco y el escenario actual de la Ciencia Abierta.

### **PALABRAS CLAVE**

Ciencia Abierta; taxonomía; comunicación científica; representación del conocimiento, Unesco.

XII Conferencia Internacional BIREDIAL-ISTEC, Uruguay, 18-20 octubre 2023

# Taxonomía de la ciencia abierta: revisada y ampliada

Consulta el artículo en:



**Lúcia da Silveira**  
Universidade Federal de Santa Catarina - UFSC  
<https://orcid.org/0000-0003-1118-2121>

**Andrea Mora-Campos**  
Universidad Nacional  
<https://orcid.org/0000-0001-9813-2674>

**Priscila Machado Borges Sena**  
Universidade Federal do Rio Grande do Sul  
<https://orcid.org/0000-0002-5612-4315>

**Fabiano Couto Corrêa da Silva**  
Universidade Federal do Rio Grande do Sul  
<https://orcid.org/0000-0001-5014-8853>

**Juliana Fachin**  
Grupo de Investigación Ciência Aberta Latam  
<https://orcid.org/0000-0003-0883-642X>

**Nivaldo Calixto Ribeiro**  
Universidade Federal de Minas Gerais  
<https://orcid.org/0000-0003-0650-0121>

**Daniel Fernando Piraquive-Piraquive**  
Universidad Distrital Francisco José de Caldas  
<https://orcid.org/0000-0002-7459-7482>

**Jorge Polanco Cortés**  
Universidad de Costa Rica  
<https://orcid.org/0000-0002-2974-5167>

**Ronaldo Ferreira Araújo**  
Universidade Federal de Alagoas  
<https://orcid.org/0000-0003-0778-9561>

**Remedios Melero**  
Instituto de Agroquímica y Tecnología de Alimentos  
<https://orcid.org/0000-0002-1813-8783>

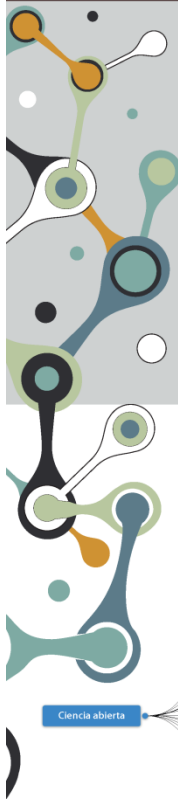
**Alejandro Uribe Tirado**  
Universidad de Antioquia  
<https://orcid.org/0000-0002-0381-1269>

**Julio Santillán-Aldana**  
Universidad Tecnológica del Perú  
<https://orcid.org/0000-0002-1906-2370>

**Andrés Mauricio Enciso Betancourt**  
Universidad Distrital Francisco José de Caldas  
<https://orcid.org/0000-0002-8348-1625>

**Palabras clave:** Ciencia Abierta - Taxonomía, Comunicación Científica Representación del Conocimiento, Unesco.  
**Keywords:** Open Science - taxonomy, Scientific communication, Representation of knowledge, Unesco.

**Eje temático:** Comunicación académica y científica en abierto



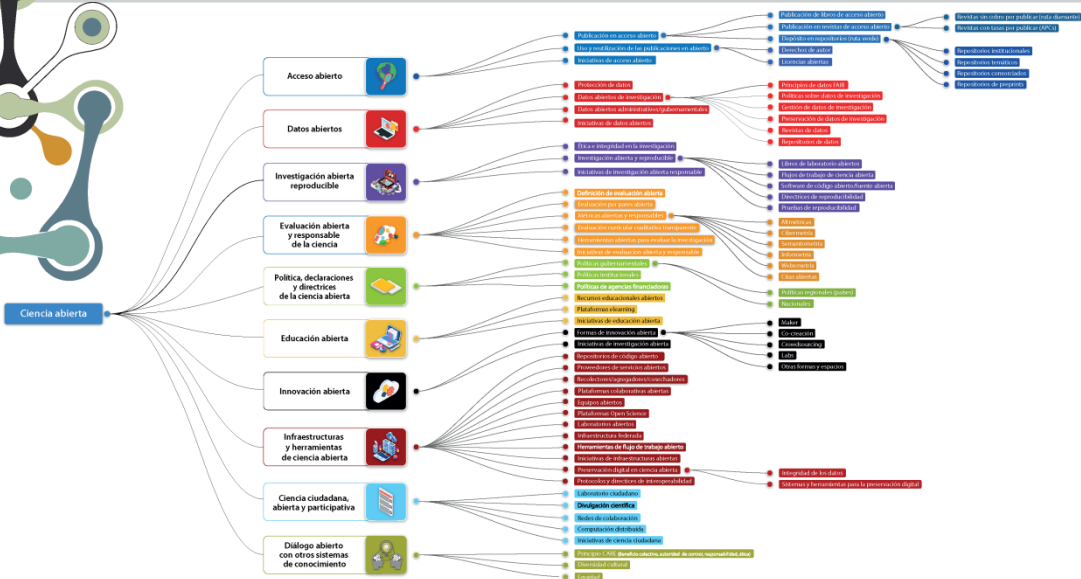
## Resumen

**Objetivo:** revisar las terminologías y aplicaciones de la taxonomía de la Ciencia Abierta, para construir una versión más completa, que represente el conocimiento en torno al tema, de acuerdo con el escenario actual de la comunicación científica y las recomendaciones de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (Unesco).

**Método:** se trata de una investigación exploratoria con un enfoque deductivo. El primer paso fue la revisión de las taxonomías, con las 12 personas dedicadas a la investigación, que se reunieron semanalmente, para las discusiones conceptuales y epistemológicas relacionadas con la Ciencia Abierta y las definiciones metodológicas y de procedimiento para la realización del estudio.

**Resultados:** como resultado de los análisis, se elaboró una taxonomía para ser evaluada por las personas expertas. Para ello, se envió a 68 personas expertas un cuestionario con preguntas abiertas sobre cada eje principal de la taxonomía. Se obtuvieron 21 respuestas que cooperaron con la modelización y exposición de los términos de la nueva propuesta de taxonomía. La nueva taxonomía tiene 10 componentes de nivel principal y un total de 96 etiquetas.

**Conclusiones:** la percepción de los expertos puso de manifiesto un panorama congruente con las recomendaciones de la Unesco y el escenario actual de la Ciencia Abierta.



Creative Commons Atribución 4.0 Internacional



ISBN 978-950-34-2375-2



9 789503 423752