



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA – FACULTAD DE LETRAS – BIBLIOTECAS UC

**IMPLEMENTACIÓN Y GESTIÓN DE UN
REPOSITORIO DE DATOS DE INVESTIGACIÓN PARA BIBLIOTECAS:
BUENAS PRÁCTICAS Y DESAFÍOS.**

AMELIA MUÑOZ BRAVO

Proyecto de Tesis para optar al Grado de
Magíster en Procesamiento y Gestión de la Información

Profesor Supervisor:
MARIBEL ALVARADO ACUÑA

Santiago de Chile, diciembre de 2020



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA – FACULTAD DE LETRAS – BIBLIOTECAS UC

**IMPLEMENTACIÓN Y GESTIÓN DE UN
REPOSITORIO DE DATOS DE INVESTIGACIÓN PARA BIBLIOTECAS:
BUENAS PRÁCTICAS Y DESAFÍOS.**

AMELIA MUÑOZ BRAVO

Proyecto de Tesis presentado a la Comisión Integrada por:

**JULIÁN SAN MIGUEL E.
ELVIRA SAURINA DE SOLANES
MARIBEL ALVARADO ACUÑA**

Santiago de Chile, diciembre de 2020

Dedicatoria

A quienes me instaron a continuar perseverando, pese a los obstáculos que encontré en este fascinante viaje de aprendizaje y conocimiento, recordándome que cada avance fue una meta por sí sola.

A quienes consagran su vida al aprendizaje constante, no como un alimento para el ego, sino como un medio para mejorar cada día y contribuir a nuestra sociedad.

A mi Familia, a mi madre (quien desde niña me incentivó a que, estudiando, lograría también progresar), a mi padre (quien me enseñó a hacer las cosas lo mejor posible; aunque partiste hace tiempo, nunca has dejado de estar en mi recuerdo), a mis hermanos y con especial afecto a Tamara Belén y a Nacho.

Agradezco a Dios, Quien siempre siento me acompaña sin preguntas.

Como escribió el científico Isaac Newton en una misiva fechada en 1675 y dirigida a Robert Hooke, anticipando hace siglos el impacto de compartir el conocimiento:

“Si he visto más lejos es porque estoy sentado sobre los hombros de gigantes”

(“If I have seen further it is by standing on the shoulders of Giants”)

Agradecimientos

A Evelyn Didier Carrasco, Denise Depoortere Piriz y a Comité de Capacitación y Desarrollo de Bibliotecas UC.

A mi profesora guía, Maribel Alvarado Acuña.

A María Elvira Saurina de Solanes.

A quienes fueron mis profesores de MPGI (Promoción 2018 – 2019).

A Mauricio Arriagada Benítez.

A Antonio Muñoz Bravo.

A Juan Ramírez Jardúa.

A Agustín Fernández-Eguiarte.

A Andrew Mckenna-Foster.

A Washington Luís R. de Carvalho Segundo.

A quienes dedicaron tiempo en ayudarme en algún momento.

Tabla de Contenidos

Dedicatoria	1
Agradecimientos	2
Tabla de Contenidos.....	3
Índice de Figuras	6
Índice de Gráficos.....	7
Índice de Tablas.....	8
Resumen.....	9
Abstract	12
1. Introducción	15
2. Motivación	17
3. Estado del arte	18
3.1. Presentación Ciencia Abierta y Acceso Abierto en Europa	19
3.1.1. Europa Acceso Abierto	19
3.1.2. Horizonte2020 Mandato de Acceso Abierto.....	20
3.1.3. ORD Pilot	24
3.1.4. B2SHARE	24
3.2. Presentación Ciencia Abierta y Acceso Abierto en Estados Unidos.....	27
3.2.1. Agencias Federales – Institutos Nacionales de Salud (NIH) como ejemplo	29
3.3. Presentación Ciencia Abierta y Acceso Abierto en América Latina	30
3.3.1. Latindex	31
3.3.2. SciELO	32
3.3.3. AmeliCA	33
3.3.4. Redalyc	34
3.3.5. LA Referencia.....	35
3.4. Presentación Ciencia Abierta y Acceso Abierto en Chile	36
3.5. Softwares para repositorios de datos de investigación	38
3.5.1. DSPACE	38
3.5.2. Dataverse	40
3.5.3. CKAN.....	42
3.5.4. INVENIO.....	43
3.5.5. FEDORA	45
3.5.6. DSPACE CRIS	46
3.6. Principios FAIR.....	47

3.6.1.	¿Por qué son importantes los Principios FAIR?	48
3.6.2.	¿Qué inspira el surgimiento de estos Principios?	49
3.7.	Certificaciones	52
3.8.	Repositorios de Próxima Generación.....	56
3.9.	Tecnologías.....	58
3.10.	Repositorios de datos de investigación en Europa	59
3.11.	Repositorios de datos de investigación en Estados Unidos	65
3.12.	Repositorios de datos de investigación en América Latina - LA Referencia	67
3.13.	Repositorios de datos de investigación en Chile.....	70
4.	Alcances y limitaciones	72
4.1.	Alcances.....	72
4.2.	Limitaciones	73
5.	Objetivos	75
5.1.	General.....	75
5.2.	Específicos	75
6.	Situación actual	76
6.1.	Unión Europea.....	77
6.2.	América Latina.....	84
6.3.	Brasil.....	87
6.4.	Chile.....	92
7.	Metodología	97
7.1.	Revisión bibliográfica	97
7.2.	Encuesta	99
7.3.	Entrevista	101
7.4.	Comunicación y colaboración digital	102
7.5.	Análisis de los casos de estudios	103
7.6.	Benchmarking de sitios web	104
8.	Aplicación a casos de estudio.....	106
8.1.	Caso 1: REPOSITORIO INSTITUCIONAL DEL CENTRO DE CIENCIAS DE LA ATMÓSFERA (UNAM)	106
8.1.1.	El origen.....	106
8.1.2.	Datos de investigación: un gran tema para los investigadores.....	108
8.1.3.	Detalle del Repositorio Institucional del Centro de Ciencias de la Atmósfera.....	111
8.1.4.	Visión futura	114

8.2.	Caso 2: FIGSHARE	115
8.2.1.	Origen y expansión.....	115
8.2.2.	Características destacadas	116
8.2.3.	Usuarios.....	118
8.2.4.	Visión futura	119
8.2.5.	Un caso exitoso de aplicación de Figshare: Universidad Estatal de Iowa	121
9.	Resultados y análisis.....	130
9.1.	Análisis de los resultados	134
9.2.	Resumen de buenas prácticas, recomendaciones y/o desafíos	160
10.	Conclusiones.....	167
	Bibliografía consultada.....	171
	Anexo 1: Exención de Responsabilidad.....	179
	Anexo 2: Formulario de Divulgación de Conjunto de Datos	181
	Anexo 3: Encuesta - Repositorios de Datos de Investigación en Bibliotecas.....	184

Índice de Figuras

Figura 3-1: Pasos Mandato HORIZONTE 2020	21
Figura 3-2: Rutas de investigación en Acceso Abierto	23
Figura 3-3: Diagrama del repositorio de datos de investigación B2SHARE.....	25
Figura 3-4: Portal latindex.....	31
Figura 3-5: Portal SciELO	32
Figura 3-6: Portal AmeliCA	33
Figura 3-7: Portal Redalyc	34
Figura 3-8: Portal LA Referencia.....	35
Figura 3-9: Ejes en Chile sobre los que se sustenta el Acceso Abierto	37
Figura 3-10: Principios FAIR.....	48
Figura 3-11: Sitio web de los Principios FAIR	52
Figura 3-12: Repositorios actuales vs Repositorios de Próxima Generación	56
Figura 3-13: Ciclo de AA y los datos de investigación almacenados en repositorios de datos.....	62
Figura 6-1: LA Referencia en América Latina	85
Figura 6-2: Modelo de agregación de LA Referencia	86
Figura 6-3: Distribución geográfica de publicaciones y datos en Brasil	88
Figura 6-4: Producción de investigación por regiones en Brasil	88
Figura 6-5: Razones por las cuales Brasil eligió a Dataverse	89
Figura 6-6: Modelos de redes de repositorios	90
Figura 6-7: Modelo de cosecha de repositorios de datos de investigación en Brasil	91
Figura 6-8: “Respecto del objetivo de esta política ¿Qué tan de acuerdo o en desacuerdo está usted?”	94
Figura 6-9: “Nivel de acuerdo/desacuerdo respecto al objetivo de la política”	94
Figura 6-10: “Indique su nivel de acuerdo o desacuerdo respecto a las disposiciones sobre datos de investigación que señala la Política de Acceso Abierto”	95
Figura 6-11: “Nivel de acuerdo/desacuerdo respecto a las disposiciones sobre datos de investigación de la política”	95
Figura 8-1: Home y temáticas del repositorio de la EEA.....	113
Figura 8-2: Home del Repositorio Institucional de Ciencias de la Atmósfera en la UNAM, con una interfaz inspirada en el repositorio de la EEA utilizando además tecnologías similares	114
Figura 8-3: Header del home principal.....	123
Figura 8-4: Home del repositorio de datos de la Universidad de Iowa.....	124
Figura 8-5: Ejemplo de registro de un conjunto de datos.....	125
Figura 8-6: Ejemplo de una categoría	126
Figura 8-7: Ejemplo de una categorías y subcategorías asociadas	126
Figura 8-8: Vista de grupos existentes	127
Figura 8-9: Ejemplo de la sección Buscar	129
Figura 8-10: Detalle de la puntuación, citación en Almetric.....	129

Índice de Gráficos

Gráfico 9-1: Total envíos por encuesta	130
Gráfico 9-2: Primer envío según zonas geográficas	131
Gráfico 9-3: Segundo envío según zonas geográficas	131
Gráfico 9-4: Total de envíos según zonas geográficas	132
Gráfico 9-5: Total de respuestas según zonas geográficas	134
Gráfico 9-6: Dotación de personal	135
Gráfico 9-7: Dotación de personal	135
Gráfico 9-8: Softwares más usados	137
Gráfico 9-9: Criterios de selección	138
Gráfico 9-10: Criterios de selección	139
Gráfico 9-11: Repositorios de datos.....	140
Gráfico 9-12: Repositorios de datos.....	140
Gráfico 9-13: Secciones para datos y otras	141
Gráfico 9-14: Secciones para datos y otras	141
Gráfico 9-15: Antigüedad de repositorios.....	142
Gráfico 9-16: Antigüedad de repositorios.....	142
Gráfico 9-17: Tipo de financiamiento.....	143
Gráfico 9-18: Tipo de financiamiento.....	143
Gráfico 9-19: Satisfacción con el software	144
Gráfico 9-20: Satisfacción con el software	145
Gráfico 9-21: Opiniones sobre experiencia con software	146
Gráfico 9-22: Seguridad de datos.....	147
Gráfico 9-23: Seguridad de datos.....	147
Gráfico 9-24: Opiniones sobre seguridad de datos.....	148
Gráfico 9-25: Satisfacción con confiabilidad	149
Gráfico 9-26: Satisfacción con confiabilidad	149
Gráfico 9-27: Opiniones sobre satisfacción.....	150
Gráfico 9-28: Aspectos complejos en implementación de software	151

Índice de Tablas

Tabla 3-1: Principios rectores FAIR.....	50
Tabla 3-2: Resumen de los tres niveles y la diferencia entre sus certificaciones	55
Tabla 3-3: Ejemplos de repositorios recomendados por Nature	67
Tabla 7-1: Resumen encuesta	100
Tabla 9-1: Aspectos complejos en implementación de software	151
Tabla 9-2: Recomendaciones y buenas prácticas.....	155
Tabla 9-3: Principales buenas prácticas, recomendaciones y/o desafíos	161

Resumen

Repositorios de Datos de Investigación para Bibliotecas es un extracto significativo del tema que presento en los capítulos, abordando desde su desarrollo en algunos países representativos en esta materia en Europa, incluyendo a Reino Unido, Australia, Estados Unidos, América Latina y Chile. Me permito proporcionar una realidad acerca del estado del arte en cada uno de ellos y comparto, además, significativas buenas prácticas y recomendaciones, emanadas de las diversas herramientas constitutivas de la metodología empleada en mi investigación.

Es así como en el capítulo **Estado del arte** presento los aspectos más recientes de los repositorios de datos de investigación, tanto desde la perspectiva de la infraestructura tecnológica como del contexto actual en que se encuentran la Ciencia Abierta y el Acceso Abierto, las cuales desempeñan un rol relevante en esta temática. También entrego definiciones y alcances de conceptos, al igual que antecedentes de organizaciones destacadas y bien posicionadas a nivel internacional en el tópico de los repositorios de datos de investigación. Por otro lado, y como consecuencia de lo anterior, concedo una mirada acerca de los grados de avance y progreso alcanzado por dichas organizaciones, subrayando el entorno favorable en el cual algunas se inscriben y destacando los esfuerzos que se realizan para tratar de nivelar el grado de desarrollo en los países que se han ido sumando a esta situación mundial. Es por esto que me refiero a los temas tecnológicos, organizacionales, legislativos, económicos y también culturales. Se describen diversas iniciativas como el programa Horizonte 2020, impulsado vigorosamente por la Unión Europea. Asimismo, se consideran proyectos en Estados Unidos (Agencias Federales de los Institutos Nacionales de Salud), al igual que en América Latina y Chile (Latindex, SciELO, AmeliCA, Redalyc, LA Referencia y la Agencia Nacional de Investigación y Desarrollo, ANID). Asimismo, presento softwares de mayor uso en los repositorios de datos de investigación, según los resultados de la metodología (DSpace, Dataverse, CKAN, INVENIO, FEDORA, DSpaceCRIS, Figshare). Luego, me centro en los llamados Principios FAIR, sin duda la columna vertebral en torno a la cual se articula todo el engranaje asociado a los repositorios de datos de investigación, señalando su significado, trascendencia y certificaciones, para dar paso a los repositorios llamados “de Próxima Generación”.

Más adelante, en **Alcances y limitaciones**, planteo la pregunta central de mi investigación relacionada con los beneficios y los desafíos que para las bibliotecas implican la implementación y gestión de un repositorio de datos de investigación, donde la metodología es fundamental para proporcionar las respuestas, como se verá más adelante (revisión bibliográfica, encuesta, entrevista, comunicación y colaboración digital, análisis de dos casos representativos de estudio y benchmarking de los sitios webs especializados). Todos estos contenidos se abordan por separado y en detalle en otros capítulos. En relación a las limitaciones inherentes a este estudio, planteo que el área de interés se circunscribe a la entrega de soluciones tecnológicas. En definitiva, dejo de manifiesto que mi tesis pretende aportar con información práctica para las instituciones que

cuentan con repositorios de datos de investigación y requieren de mejoras o pretenden implementarlos.

Respecto de los **Objetivos**, señalo como norte la recopilación y la entrega tanto de pautas como de recomendaciones para la implementación, gestión y mejora continua de un repositorio de datos de investigación. En cuanto a los objetivos específicos, analizo una muestra relevante de los repositorios de datos existentes en el mundo con el fin de conocer y de comprender sus principales características, presento el estado de la Ciencia Abierta y el Acceso Abierto y sus comportamientos en distintos continentes, identifico los recursos y las buenas prácticas utilizadas en los repositorios de datos de investigación -como lo he mencionado- y describo la situación en Chile respecto de este tema.

Luego, en **Situación actual**, entrego una fotografía de la realidad en que se encuentran los 34 países de la Unión Europea adscritos a la infraestructura tecnológica y de servicios conocida como OpenAIRE, señalando -entre otros- los avances alcanzados en materia de funcionamiento, estándares de calidad, cooperación y legislación. Refiero la evolución histórica de las soluciones tecnológicas y aludo a los desafíos futuros. El mismo análisis efectúo en América Latina, donde el caso de la plataforma LA Referencia se ha perfilado como el más entusiasta seguidor de las propuestas sobre repositorios de datos de investigación formuladas desde Europa, a la vanguardia indiscutible en estas materias. A continuación, analizo el caso de Brasil, país que ha avanzado mucho en comparación a otros de la región, con una trayectoria de larga data y alineado ciento por ciento con LA Referencia, estimulando el desarrollo de las investigaciones. Por último, me refiero a nuestro país, donde la ANID se encuentra impulsando iniciativas para posicionar a Chile dentro de la comunidad internacional, al alero de LA Referencia, de la cual forma parte. Incluyo los resultados de una encuesta lanzada por este organismo estatal entre junio y julio de 2020, teniendo como objetivo conocer la postura de los diversos actores relacionados con la nueva política de Acceso Abierto que se pretende implementar en el futuro y cuyos resultados ya fueron presentados en diciembre de 2020.

En cuanto a la **Metodología**, me apoyo en seis recursos e instrumentos. Uno de ellos es la revisión bibliográfica, con fuentes de información en Acceso Abierto en un período comprendido entre 2017 y diciembre de 2020, con cobertura idiomática en español y en inglés, abarcando el estado del arte en estas materias en Europa, Estados Unidos, Canadá, Australia, Latinoamérica y Chile. Se incluyeron conceptos y palabras clave, a partir de las cuales tuvieron lugar nuevas búsquedas relacionadas. También se generaron “alertas” vinculadas al tema de interés en los recursos bibliográficos digitales revisados, con el propósito de acceder a lo más reciente. El siguiente instrumento fue la encuesta que desarrollé con el propósito de hacer un levantamiento de contenidos sobre los repositorios de datos, al igual que de recomendaciones y de buenas prácticas, con dos envíos (octubre y noviembre) y englobando un total de 335 organizaciones. En cuanto a la entrevista, incluyo un diálogo valioso con el investigador y destacado académico Agustín Fernández Eguiarte, de la Universidad Nacional Autónoma de México (UNAM), pionero en la generación de un repositorio de datos en dicha institución académica. Luego, en el apartado de “Comunicación y colaboración digital”, expongo los

aportes que recibí de parte de Andrew McKenna-Foster, de Figshare, y de Washington Segundo, Coordinador del Laboratorio de Metodologías de Tratamiento y Difusión de Información (COLAB), del Instituto Brasileño de Información en Ciencia y Tecnología (IBICT / MCTIC). Acto seguido, formulo un análisis de dos casos de estudio exitosos y que tributan a los objetivos planteados en esta publicación. Al final, incluyo el benchmarking de los sitios web especializados, estableciendo puntos de concordancia.

En **Aplicación a casos de estudios** abordo, en primer lugar, la experiencia del Repositorio Institucional del Centro de Ciencias de la Atmósfera (UNAM), refiriéndome a sus orígenes, conformación del Repositorio y administración de los datos, así como a las tecnologías empleadas en su desarrollo, las características principales, la interfaz web y la visión futura, considerando también aspectos humanos, de gestión, de recursos y administrativos. En segundo lugar, considero la plataforma Figshare, una de las de mayor presencia en el mundo, entregando información relacionada con su historia y evolución, características singulares, usuarios o clientes en el mundo y visión futura. Para reforzar el análisis, ilustro con uno de los muchos casos exitosos de aplicación práctica, como es el repositorio de datos de investigación de la Universidad Estatal de Iowa.

A continuación, paso a los **Resultados y análisis**, construido sobre los resultados obtenidos a partir de la encuesta titulada “Repositorios de Datos de Investigación en Biblioteca”, de mi autoría y gestionada a través de la plataforma LibWizard de Springshare. Teniendo como base 15 preguntas en español e inglés, de selección y abiertas, tuvo como objetivo conocer las experiencias en la implementación y desarrollo de un repositorio de datos en diversas instituciones del mundo (Europa, Reino Unido, Estados Unidos, Canadá, Australia y América Latina), para lo cual se consultó sobre temas como software, personal, antigüedad del repositorio y grado de satisfacción, entre otros. El análisis se complementa con gráficos de barras y de mesa. Por separado, me refiero a las recomendaciones y a las buenas prácticas recogidas no sólo de las dos últimas preguntas de la mencionada encuesta, sino también de los otros recursos e instrumentos que conforman la metodología del presente estudio.

Por último, entrego las **Conclusiones** que surgen luego de una concienzuda reflexión, reforzando algunas de las respuestas a las interrogantes planteadas al inicio, las cuales fui contestando a lo largo de los capítulos previos. A modo de síntesis, sostengo que una de las principales recomendaciones para incursionar en la implementación de un repositorio de datos de investigación es reflexionar, previamente, sobre las necesidades estratégicas de una organización. También enfatizo la necesidad de que las organizaciones académicas internalicen las sugerencias y directrices que -en Ciencia Abierta, Acceso Abierto y repositorios de datos- emanan desde Europa, por lejos a la vanguardia, con el fin de contribuir al conocimiento teórico y práctico, integrándose de paso a la comunidad internacional. Esto implica una participación activa de los investigadores, parte constitutiva del espíritu de una biblioteca al igual que su personal. Concuerdo con el fomento de los Principios FAIR, con la importancia de los recursos tecnológicos, económicos y humanos, y con el estímulo constante de una cultura que propicie la adhesión y permanencia en esta comunidad.

Abstract

Research Data Repositories for Libraries is a significant extract from the topic that I present in the chapters, addressing from its development in some representative countries in this matter in Europe, including the United Kingdom, Australia, United States, Latin America and Chile. I allow myself to provide a reality about the state of the art in each one of them and I also share significant good practices and recommendations, emanating from the various constitutive tools of the methodology used in my research.

This is how in the chapter **State of the art**, I present the most recent aspects of the research data repositories, both from the perspective of technological infrastructure and the current context in which Open Science and Open Access are found, which play an important role. relevant role in this issue. I also provided definitions and scopes of concepts, as well as the background of prominent and well-positioned organizations at an international level on the topic of research data repositories. On the other hand, and as a consequence of the foregoing, I give a look at the degrees of advancement and progress achieved by these organizations, highlighting the favorable environment in which some are enrolled and highlighting the efforts made to try to level the grade of development in the countries that have been added to this world situation. This is why I refer to technological, organizational, legislative, economic and also cultural issues. Various initiatives are described, such as the Horizon 2020 program, vigorously promoted by the European Union. Likewise, projects are considered in the United States (Federal Agencies of the National Institutes of Health), as well as in Latin America and Chile (Latindex, SciELO, AmeliCA, Redalyc, LA Reference and the National Research and Development Agency, ANID). Likewise, I present the most widely used software in research data repositories, according to the results of the methodology (DSpace, Dataverse, CKAN, INVENIO, FEDORA, DSpaceCRIS, Figshare). Then, I focus on the so-called FAIR Principles, undoubtedly the backbone around which all the gear associated with research data repositories is articulated, pointing out their meaning, significance and certifications, to give way to the repositories called “Next Generation”.

Later, in **Scopes and limitations**, I pose the central question of my research related to the benefits and challenges that the implementation and management of a research data repository imply for libraries, where the methodology is fundamental to provide the answers, such as It will be seen later (bibliographic review, survey, interview, communication and digital collaboration, analysis of two representative case studies and benchmarking of specialized websites). All these contents are dealt with separately and in detail in other chapters. In relation to the limitations inherent to this study, I suggest that the area of interest is limited to the delivery of technological solutions. In short, I make it clear that my thesis aims to provide practical information for institutions that have repositories of research data and require improvements or intend to implement them.

Regarding the **Objectives**, I point out as north the compilation and delivery of both guidelines and recommendations for the implementation, management and continuous improvement of a

repository of research data. Regarding the specific objectives, I analyze a relevant sample of the existing data repositories in the world in order to know and understand their main characteristics, I present the state of Open Science and Open Access and their behaviors in different continents, I identify the resources and good practices used in the research data repositories -as I have mentioned- and I describe the situation in Chile regarding this issue.

Then, in **Current situation**, I give a photograph of the reality in which the 34 countries of the European Union are attached to the technological and service infrastructure known as OpenAIRE, pointing out -among others- the progress made in terms of operation, standards of quality, cooperation and legislation. I refer to the historical evolution of technological solutions and refer to future challenges. The same analysis was carried out in Latin America, where the case of the LA Reference platform has emerged as the most enthusiastic follower of the proposals on research data repositories made from Europe, at the undisputed vanguard in these matters. Next, I analyze the case of Brazil, a country that has advanced a lot compared to others in the region, with a long history and aligned one hundred percent with LA Reference, stimulating the development of research. Lastly, I am referring to our country, where ANID is promoting initiatives to position Chile within the international community, under the wing of LA Reference, of which it is a part. I include the results of a survey launched by this state body between June and July 2020, with the objective of knowing the position of the various actors related to the new Open Access policy that is intended to be implemented in the future and whose results have already been presented in December 2020.

Regarding the **Methodology**, I rely on six resources and instruments. One of them is the bibliographic review, with Open Access information sources in a period between 2017 and December 2020, with language coverage in Spanish and English, covering the state of the art in these matters in Europe, the United States, Canada, Australia, Latin America and Chile. Concepts and keywords were included, from which new related searches took place. "Alerts" were also generated related to the topic of interest in the reviewed digital bibliographic resources, in order to access the most recent. The next instrument was the survey that I developed with the purpose of collecting content on data repositories, as well as recommendations and good practices, with two shipments (October and November) and encompassing a total of 335 organizations. Regarding the interview, I include a valuable dialogue with the researcher and prominent academic Agustín Fernández Eguiarte, from the National Autonomous University of Mexico (UNAM), a pioneer in the generation of a data repository at said academic institution. Then, in the section on "Communication and digital collaboration", I present the contributions I received from Andrew McKenna-Foster, from Figshare, and from Washington Segundo, Coordinator of the Laboratory of Information Treatment and Dissemination Methodologies (COLAB), of the Brazilian Institute of Information in Science and Technology (IBICT / MCTIC). Next, I formulate an analysis of two successful case studies that contribute to the objectives set out in this publication. In the end, I include the benchmarking of specialized websites, establishing points of agreement.

In **Application to case studies on board**, first of all, the experience of the Institutional Repository of the Center for Atmospheric Sciences (UNAM), referring to its origins, composition of the Repository

and data management, as well as the technologies used in its development, the main characteristics, the web interface and the future vision, also considering human, management, resource and administrative aspects. Secondly, I consider the Figshare platform to be one of the most present in the world, providing information related to its history and evolution, unique characteristics, users or clients in the world, and future vision. To reinforce the analysis, I illustrate with one of the many practical application success stories, such as the Iowa State University Research Data Repository.

Next, I move on to the **Results and analysis**, built on the results obtained from the survey entitled "Repositories of Research Data in Library", of my authorship and managed through the Springshare LibWizard platform. Based on 15 questions in Spanish and English, selection and open, the objective was to learn about the experiences in the implementation and development of a data repository in various institutions around the world (Europe, United Kingdom, United States, Canada, Australia and America Latina), for which it was consulted on issues such as software, personnel, age of the repository and degree of satisfaction, among others. The analysis is complemented by bar and table graphs. Separately, I refer to the recommendations and good practices collected not only from the last two questions of the aforementioned survey, but also from the other resources and instruments that make up the methodology of this study.

Finally, I deliver the **Conclusions** that emerge after a thorough reflection, reinforcing some of the answers to the questions raised at the beginning, which I was answering throughout the previous chapters. By way of synthesis, I argue that one of the main recommendations to venture into the implementation of a research data repository is to reflect, previously, on the strategic needs of an organization. I also emphasize the need for academic organizations to internalize the suggestions and guidelines that - in Open Science, Open Access and data repositories - emanate from Europe, far at the forefront, in order to contribute to theoretical and practical knowledge, integrating from step to the international community. This implies the active participation of researchers, a constitutive part of the spirit of a library as well as its staff. I agree with the promotion of the FAIR Principles, with the importance of technological, economic and human resources, and with the constant encouragement of a culture that encourages adherence and permanence in this community.

1. Introducción

El presente trabajo se inserta dentro del Seminario de Graduación del Magíster en Procesamiento y Gestión de Información (MPGI) 2018-2019, impartido conjuntamente por la Escuela de Ingeniería, la Facultad de Letras y Bibliotecas UC, de la Pontificia Universidad Católica de Chile (PUC).

El título de la publicación es “Implementación y Gestión de un Repositorio de Datos de Investigación para Bibliotecas: Buenas Prácticas y Desafíos”. Corresponde a una revisión que busca ser un aporte en especial para las bibliotecas que requieren desarrollar, mejorar o gestionar infraestructura tecnológica adecuada, como es un repositorio de datos, recurso que acompaña el ciclo de vida de la gestión de datos de investigación¹ y cuyo propósito es facilitar a la producción académica y de investigación un espacio digital para el depósito de los resultados obtenidos, los cuales pueden estar en variados formatos, desde hojas de cálculo, estadísticas, resultados de encuestas, entrevistas, resultados de experimentos, cuadernos de laboratorio, fotografías en alta resolución y grabaciones, entre otros, permitiendo la accesibilidad de quienes lo necesiten y/o el intercambio y la colaboración entre sus pares.

La Ciencia Abierta (CA), incipiente aún en nuestra realidad nacional, corresponde a un movimiento originado en Europa cuyo objetivo es hacer que la creación científica sea accesible por todas las instancias de la sociedad, impulsando que todos los resultados de la investigación financiados con fondos públicos o estatales sean difundidos o accesibles para cualquiera que así lo requiera. Para lograr lo anterior, la CA articuló, entre muchas iniciativas, algunos principios como metodologías e infraestructuras tecnológicas: softwares y hardware, revisiones por pares, recursos educativos abiertos (MOOC) y datos. El Acceso Abierto (AA) es la acción inmediata, sin ninguna restricción, de entregar los resultados, siendo éste el caso de los datos de investigación depositados en repositorios. Ambas instancias también se incluyen en este trabajo, así como una panorámica de su estado actual.

En este escenario surge la necesidad de disponer de herramientas o medios tecnológicos capaces de acopiar, preservar, entregar y/o compartir -con quienes lo requieran- estos resultados para el caso presentado: datos de investigación.

Para hacer posible lo anterior, se incluye en este trabajo información de buenas prácticas y recomendaciones vigentes para la implementación de este tipo de repositorios, estado del arte, una descripción de la situación nacional e internacional de la temática, entre otros elementos que se consideran relevantes al momento de crear o aplicar mejoras en un repositorio de datos.

¹ Research Data Management (RDM). Administración efectiva de la información que se crea en el curso de la investigación. Más datos en: <https://www.jisc.ac.uk/guides/how-and-why-you-should-manage-your-research-data> Consultado 290620.

Repositorios es el nombre con el cual se conoce a las plataformas o softwares tecnológicos que emergieron bajo el alero de centros de investigación e instituciones universitarias, incluyendo a las bibliotecas, cuyo objetivo principal fue -al inicio- almacenar y preservar investigaciones. Sin embargo, el desigual posicionamiento y regulación en materia de AA entre los continentes -y dentro de estos mismos (en países como los pertenecientes a América Latina)-, el creciente nivel de especialización en la investigación y las diferencias en relación al avance entre disciplinas, entre otros factores, se encargaron de evidenciar progresos desiguales en materia de implementación y de gestión de los repositorios. Pese a estos disímiles aspectos -entre muchos otros, tales como el desnivel de conocimiento, muchas veces producto de que esta temática no está en la agenda de las prioridades, lo cual pudo impactar en su desarrollo y en su gestión-, existe en la actualidad consenso mundial en cuanto a que los repositorios se han ido fortaleciendo con el posicionamiento de políticas orientadas al AA, convirtiendo así a los diferentes tipos de repositorios en un actor principal conocido como la Vía Verde.

Como se señaló, estas páginas versarán sobre “Repositorios de Datos”, proporcionando de manera complementaria a lo ya mencionado, información de los principales softwares destinados para su desarrollo y recomendaciones para que las bibliotecas puedan optimizar sus recursos, conociendo experiencias previas al momento de la planificación de un proyecto de esta envergadura o cuando quieran, por ejemplo, aplicar mejoras si ya disponen de un repositorio en el cual deseen incorporar más datos de investigación. Por otra parte, también se entregará información de los Repositorios de Próxima Generación.

Se reitera que el contenido de este trabajo pueda ser de utilidad principalmente para la toma de decisiones por parte de bibliotecas interesadas en esta materia, quienes pueden tener la necesidad o estar frente al desafío de emprender con la implementación y gestión de un repositorio de datos. Para ello, es imprescindible conocer aspectos estratégicos que no siempre resultan tan evidentes, optimizando y repercutiendo exitosamente en un proyecto tecnológico de este tipo, el que por cierto requiere de una mirada global que garantice su posicionamiento y escalabilidad, aspectos que también deben ser conocidos en el quehacer de los profesionales de la información.

2. Motivación

La principal motivación para abordar el tema de los Repositorios de Datos se inspira en que, actualmente, esta temática es de desarrollo incipiente en Chile. Esta posición se convierte en una oportunidad para conocer de un modo sistemático los contenidos inherentes a este tema, con el propósito de adquirir y, eventualmente, desarrollar destrezas para compartir en el ámbito de las bibliotecas, donde se percibe con mayor fuerza la irrupción de este tipo de repositorios.

La CA y el AA han impulsado esta infraestructura tecnológica denominada repositorios -y, en este caso, repositorios de datos-, lo cual genera e insta a una serie de interrogantes que son necesarias de afrontar y que pudiesen ser atendidas con el contenido de este trabajo. Estamos frente a un tema disímil en su adopción y desarrollo en el mundo, por lo que debe entender su origen conociendo a cabalidad la necesidad institucional, la visión de hasta dónde se espera llegar, entre otros aspectos que pueden ser considerados y planteados, atendiendo a interrogantes tales como ¿por qué la biblioteca debería implementar un repositorio de datos de investigación?, ¿cómo se emprende con un repositorio de datos?, ¿quiénes deberían relacionarse con la implementación y posterior gestión?, ¿existe un marco regulatorio?, ¿qué recursos se deben considerar para un proyecto de esta naturaleza?, ¿qué beneficios y desafíos implicaría para la biblioteca?, ¿qué software elegimos para la implementación?, ¿abordamos el desafío en la institución o adherimos a alguna organización que ya dispone de repositorios de datos? Ocuparse de estas preguntas u otras que surjan es motivo, también, para impulsar en las bibliotecas el ejercicio obligado de reflexionar previamente al emprendimiento de un proyecto como es un repositorio de datos, lo cual implica mucho más que sólo habilidades tecnológicas. Si se considera esto, se propiciará un impacto favorable que se traducirá en un sello diferenciador en las bibliotecas, cuya metodología para emprender con un repositorio de datos requiere atender a los aspectos señalados, sobre los cuales me referiré en la medida de lo posible dentro de este mismo documento.

Desde otra perspectiva, este tema también ha despertado mi interés debido al actual escenario a nivel nacional y a la consulta pública lanzada en mayo de este año por la Agencia Nacional de Investigación y Desarrollo (ANID) referente a una Política de Acceso Abierto a Información Científica² con el propósito de que Chile pueda adoptar una política sobre esta materia en el 2021, donde se consideren también los datos abiertos. Por ende, las bibliotecas requieren con urgencia comprender, conocer y tener una postura frente a dichas temáticas, contexto en el que la presente publicación puede considerarse como un aporte de conocimientos y de experiencias, facilitando así la toma de decisiones adecuadas o ayudando a proyectar la implementación de un repositorio institucional de datos y/o proyecciones para formar parte de una red local o regional.

² Noticia en sitio web:

<https://www.anid.cl/2020/05/27/consulta-publica-sobre-politica-de-acceso-abierto-a-informacion-cientifica-y-datos-de-investigacion/>

3. Estado del arte

En este capítulo se presentan los aspectos más recientes de los repositorios de datos de investigación, tanto desde la perspectiva de la infraestructura como del contexto actual en relación al estado en que se encuentran la CA y el AA en Europa, Estados Unidos y América Latina, áreas que definitivamente impulsan el avance en esta materia.

Aquí se encontrarán una serie de posiciones recogidas desde la literatura y originadas en variados escenarios, así como referentes teóricos, los cuales acompañan el progreso de este tipo de repositorios. Se entregarán definiciones y alcances de conceptos, al igual que antecedentes de organizaciones destacadas y bien posicionadas, las cuales representan y tratan de facilitar la unificación de tendencias políticas y tecnológicas con el objetivo de posibilitar el conocimiento, compartir experiencias traducidas en nuevas comprensiones y avances o generación de nuevos conocimientos para quienes vean en este soporte tecnológico un valor significativo el cual requiere de atención para alcanzar y formar parte activa en el ciclo de vida de la investigación interdisciplinar.

Asimismo, este panorama permitirá apreciar disímiles estados de desarrollo en el mundo, la vorágine tecnológica alcanzada por algunos repositorios, en contraposición de grandes esfuerzos que realizan otros, para propiciar la evolución de estos soportes que requieren tecnología de vanguardia y múltiples protocolos y estándares que permitan su implementación y gestión, sin dejar de considerar en esto la existencia de políticas y mandatos dictaminadas en distintos países en relación a su postura, lo cual también puede impactar.

Por Repositorio de Datos de Investigación entenderemos “un sistema de información técnica y organizativa que ayuda a los investigadores a gestionar, almacenar y proporcionar sus propios conjuntos de datos y a encontrar y acceder fácilmente a conjuntos de datos de otras fuentes” (Kindling, 2017). Por su parte, la fase de recopilación e investigación bibliográfica demostró que existen repositorios de datos de investigación de tipo generalistas y disciplinares o temáticos, lo cual se reitera como una tendencia para diferenciarse de los repositorios de datos gubernamentales. Se considera importante mencionar esta diferencia en la clasificación, cuya precisión podría ser considerada al momento de formular una adecuada definición por parte de una institución que requiera mejorar esta tecnología o tenga en perspectiva su implementación.

Asimismo, este capítulo evidenciará también que los repositorios de datos de investigación están bajo el alero de o fueron concebidos por organismos que tienen años de tradición, creación y aportes en el conocimiento disciplinar; o bien, corresponden a repositorios que son parte de bibliotecas, las cuales han sido capaces de adoptar y desarrollar conocimientos y habilidades relacionadas a esta infraestructura tecnológica, con el propósito de instaurar y mantener el ciclo de vida de la investigación.

3.1. Presentación Ciencia Abierta y Acceso Abierto en Europa

Como se señaló, la CA y el AA han sido pilares fundamentales en la compleja articulación para la creación de repositorios de datos de investigación. El posicionamiento de la CA y AA experimentan disímiles realidades, incluso estados en transición, lo cual se reflejará a continuación, con especial énfasis, en algunos países de Europa, al igual que en Estados Unidos y en América Latina, de modo particular en Chile.

3.1.1. Europa Acceso Abierto

Se presenta una panorámica general correspondiente a algunas declaraciones e iniciativas, las cuales pueden ser apreciadas como hitos, reconocidos además ampliamente en la literatura referida a AA:

- **2002: Iniciativa de Acceso Abierto de Budapest (BOAI)**³ cuyo encuentro entre varios países se originó para abordar inconvenientes y lograr acuerdos en relación a la comunicación científica afectada por las demoras del proceso editorial y el alto costo de las suscripciones de revistas, en un escenario caracterizado por el desaprovechamiento de las posibilidades ofrecidas por Internet.
- **2003: Declaración de Berlín sobre el Acceso Abierto al Conocimiento en Ciencias y Humanidades**⁴ Encuentro que suscitó una asistencia mayor que el anteriormente referido, donde se relevó y asumió que Internet tenía un impacto real en la praxis y en los aspectos económicos generadores del conocimiento y de la cultura, los cuales -además- deberían repercutir como un beneficio de acceso mundial.
- **2014-2020: Programa Horizonte 2020**⁵ Instaurado por la Comunidad Europea (CE) para impulsar y acrecentar la investigación e innovación en variadas áreas del conocimiento, tales como agricultura y bosques, biotecnología, energía, medio ambiente y acción climática, nutrición, salud, tecnologías clave de habilitación⁶, plástica, ciencias sociales y humanidades, espacio, transporte, entre otros tópicos, en los países integrantes de esta entidad con el propósito de instaurar competitividad en la investigación como un ente articulador en aspectos económicos para garantizar el desarrollo y bienestar de la sociedad europea. Este programa declaró disponer con cerca de ochenta mil millones de euros para

³ Budapest Open Access Initiative (BOAI o Iniciativa de Acceso Abierto de Budapest, en español). Sitio web: <https://www.budapestopenaccessinitiative.org/open-access-toward-the-internet-of-the-mind>

⁴ The Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. Declaración de Berlín en sitio web: <https://openaccess.mpg.de/Berlin-Declaration>

⁵ Noticia disponible en sitio oficial dependiente de la Unión Europea: <https://ec.europa.eu/programmes/horizon2020/en>

⁶ Las Tecnologías Clave de Habilidadación (o Key Enabling Technologies, KETs) son inversiones y tecnologías que permitirán a las industrias europeas conservar la competitividad y capitalizar nuevos mercados. Se centran en cuatro KET: nanotecnologías, materiales avanzados, biotecnología y desarrollo sustentable. También se investigan en edificios con energía inteligente, industrias del futuro e industrias con procesos sustentables, según sitio web: <https://ec.europa.eu/programmes/horizon2020/en/area/key-enabling-technologies>

invertir en el lapso de siete años. Fue creado para que distintas entidades, desde empresas comerciales y hasta bibliotecas, pudiesen postular proyectos en el marco de las temáticas mencionadas para lograr, además, contribuir a una ciencia de excelencia y/o liderazgo industrial, incorporando el bienestar de sus comunidades.

3.1.2. Horizonte2020 Mandato de Acceso Abierto

Horizonte2020 Mandato de AA es un programa público de gran envergadura, el cual también vinculó, a través de un mandato (2015), un soporte tecnológico para el tratamiento del AA y de los datos abiertos denominado OpenAIRE⁷ correspondiente a una entidad que colabora desde una perspectiva de infraestructura digital en todas las instancias de este precepto, abarcando un ecosistema abierto⁸ donde convergen: datos abiertos, educación abierta, gobierno abierto, licencias abiertas, educación abierta, ciencia abierta⁹ y software de código abierto¹⁰.

En el mandato, la CE dispuso que todas las publicaciones revisadas por pares y generadas en proyectos financiados por el programa Horizonte 2020 fuesen depositadas en AA, garantizando así la accesibilidad y disponibilidad de modo gratuito, sin restricciones de uso en sistemas online. Por otro lado, difundió cuatro pasos que debiesen ser observados por parte de los investigadores asociados a proyectos originados en este nuevo contexto, los cuales son presentados en la siguiente figura:

⁷ Open Access Infrastructure for Research in Europe (OpenAIRE o Infraestructura de Acceso Abierto para la Investigación en Europa, en español). Sitio web: <https://www.openaire.eu/>

⁸ El Acceso Abierto se concibe como un ecosistema donde la investigación se organiza a partir de diferentes elementos que interactúan entre sí, formando partes de un todo, según publicación comentada en página dependiente de la Universidad de Salamanca. Más datos en: <https://diarium.usal.es/gredos/2018/10/21/2700/>

⁹ La Ciencia Abierta abarca todo el ciclo de producción del conocimiento científico, incluida la concepción, la recopilación de datos, procesamiento, publicación y distribución, o bien la reutilización y la evaluación de los resultados, según artículo publicado en sitio de Agencia SINC: <https://www.agenciasinc.es/Reportajes/Movimiento-open-science-un-nuevo-plan-para-liberar-a-la-ciencia-europea>

¹⁰ Artículo incluido en sitio web: <https://www.slideshare.net/OpenAccessBelgium/20151019-webinar-open-access-in-horizon-2020>

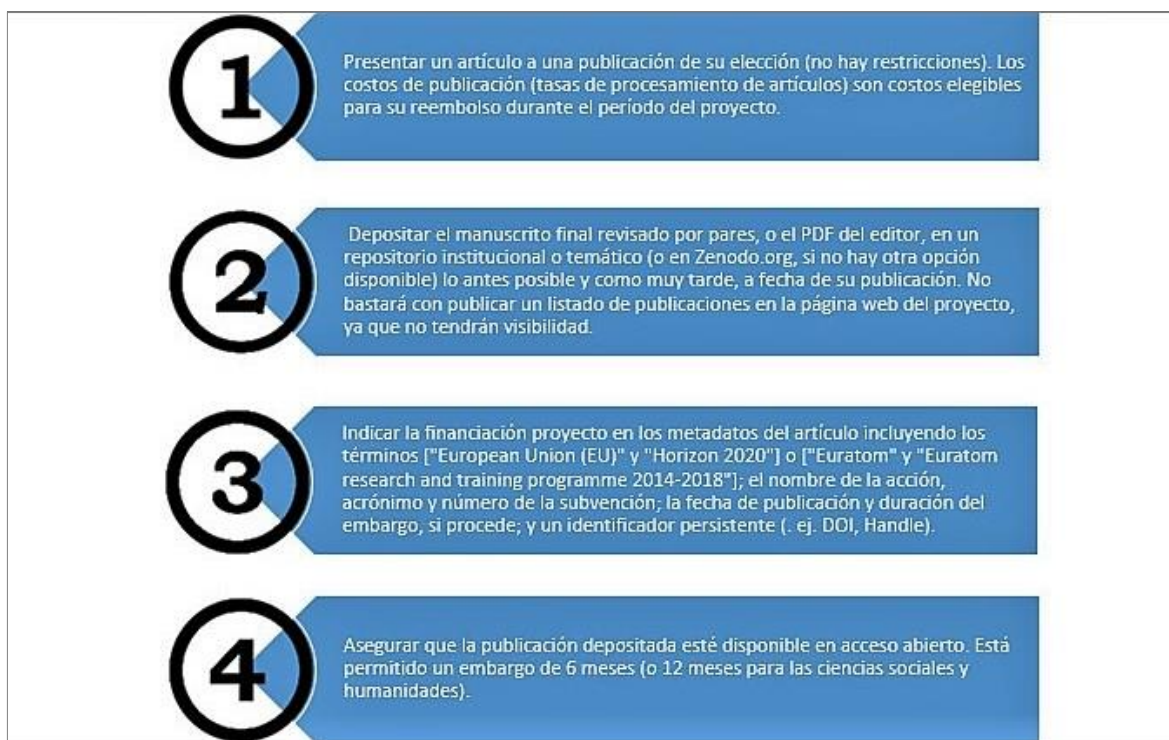


Figura 3-1: Pasos Mandato HORIZONTE 2020

Fuente: Adaptado de:

http://www2.unavarra.es/qesadj/servicioBiblioteca/piloto_de_datos/OpenAire_HojasInformativas.pdf

De acuerdo con lo expuesto anteriormente, OpenAIRE se plantea como un equipamiento consolidado y de soporte de AA, el cual se suma a la promoción de la investigación a través de servicios, prestaciones e infraestructura tecnológica, velando para que todos los formatos de documentos resultantes de investigaciones financiadas con recursos de Horizonte2020 sean depositados en repositorios que hayan adoptado sus directrices, tales como la utilización de metadatos y estándares correctos, para así alcanzar la regulación del AA, entre otros aspectos. Por otro lado, OpenAIRE tiene injerencia en la recolección de repositorios de datos y puede aplicar minería de datos sobre éstos, con el objeto de relacionar publicaciones financiadas con datos de investigación obtenidos, lo cual va propiciando el descubrimiento, la reutilización de datos, el seguimiento y la capacidad de entregar métricas, otorgando el dinamismo necesario que constituye este ciclo de gestión de datos de investigación.

En el contexto de una cooperación estratégica respecto de la gestión de datos de investigación, re3data.org¹¹ y OpenAIRE firmaron en 2013 un acuerdo de intenciones (MoU), con el objetivo de incentivar y desarrollar un potente ecosistema donde se incluían infraestructuras, las cuales permitirían el acceso permanente, el intercambio de datos o la reutilización de éstos para apoyar a quienes generaban investigaciones y conocimiento. Cabe hacer notar que esta es una de las

¹¹ re3data es un registro global de repositorios de datos de investigación que cubre repositorios de datos de investigación de diferentes disciplinas académicas y que nació en 2012. Más datos en: <https://www.re3data.org/about>

premisas requeridas por la CA que además instiga a publicar en repositorios de datos de investigación.¹²

La CE también nos aporta con una definición para el AA, refiriéndose a éste como “la práctica de proporcionar acceso en línea a información científica que es gratuito para el usuario final y reutilizable. ‘Científico’ se refiere a todas las disciplinas académicas. En el contexto de la investigación y la innovación, ‘información científica’ puede significar: artículos de investigación científica revisados por pares (publicados en revistas académicas), o datos de investigación (datos subyacentes a publicaciones, datos curados y / o datos brutos)”¹³ (CE. Comunidad Europea, s.f.).

En relación al segundo significado, y en directa referencia a esta publicación, el AA a datos de investigación se refiere “al derecho a acceder y reutilizar los datos de investigación digital en los términos y condiciones establecidos en el Acuerdo de Subvención”¹⁴ (CE. Comunidad Europea, s.f.).

Asimismo, en relación a los Datos de la Investigación, “se refiere a información, en particular hechos o cifras, recopilada para ser examinada y considerada como base para el razonamiento, la discusión o el cálculo. En un contexto de investigación, los ejemplos de datos incluyen estadísticas, resultados de experimentos, mediciones, observaciones resultantes del trabajo de campo, resultados de encuestas, grabaciones de entrevistas e imágenes. La atención se centra en los datos de investigación que están disponibles en forma digital. Los usuarios normalmente pueden acceder, extraer, explotar, reproducir y difundir datos de investigación de libre acceso de forma gratuita”¹⁵ (CE. Comunidad Europea, s.f.).

¹² Fundación Española para la Ciencia y la Tecnología/ Recolecta, Recolector de Ciencia Abierta. Sitio web: <https://recolecta.fecyt.es/node/479>

¹³ Consultado 040720.

¹⁴ Consultado 040720.

¹⁵ Consultado 040720.

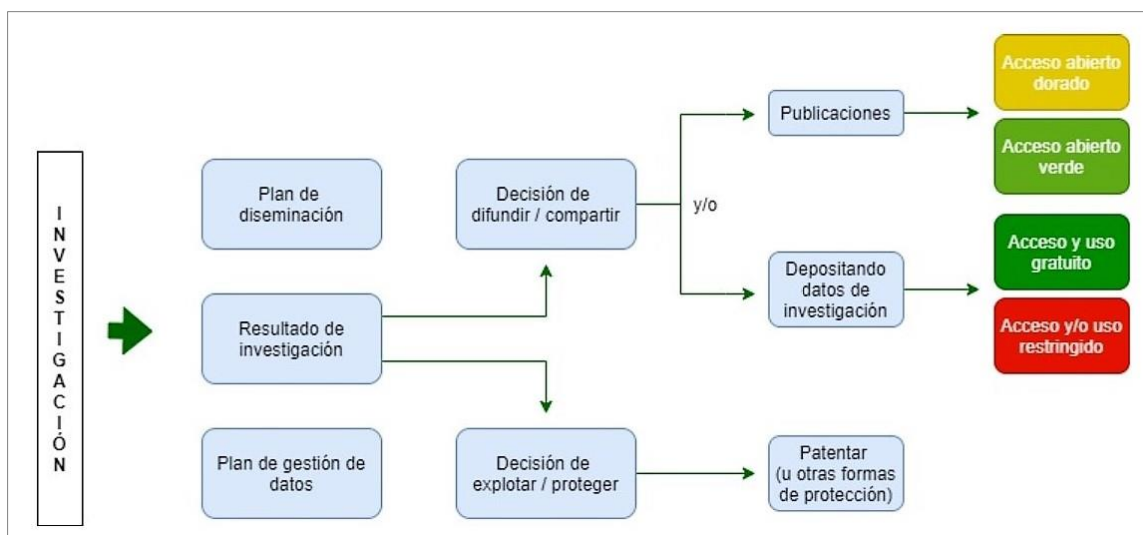


Figura 3-2: Rutas de investigación en Acceso Abierto

Fuente: Horizonte 2020 Online Manual, European Commission:

https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm

Como se aprecia en la Fig. 2, la CE materializó -bajo el programa 2020- las rutas de la investigación en AA para las publicaciones, los datos de investigación y las patentes.

En lo que se refiere a su interés por alcanzar el AA respecto no sólo de las publicaciones, sino también de los datos de investigación, el programa Horizonte2020 pretende “aprovechar los resultados de investigaciones anteriores (mejora de la calidad de los resultados), fomentar la colaboración y evitar la duplicación de esfuerzos (mayor eficiencia), acelerar la innovación (un progreso más rápido en el mercado significa un crecimiento más rápido) e involucrar a los ciudadanos y la sociedad (mayor transparencia del proceso científico)”¹⁶ (CE. Comunidad Europea, s.f.).

Este programa sostiene firmemente que toda investigación financiada con fondos públicos no debería tener un costo para acceder a dichos resultados, sin importar el ámbito desde donde surja el interés.

En 2017, en el campo de los datos de investigación, Horizon 2020 estableció un Piloto Extendido sobre AA a Datos de Investigación (ORD Pilot) para permitir el AA y la reutilización de los datos. De acuerdo con los planteamientos de ORD Pilot, todos los proyectos o investigaciones que generen datos de investigación deben cumplir con dos requisitos:

- Depositar los datos de investigación obtenidos de preferencia en un repositorio de datos de investigación (temáticos), sean generales, institucionales o de otro tipo. A modo de ejemplo,

¹⁶ Consultado 040720.

se puede citar el caso de Zenodo¹⁷, el cual -además de permitir el depósito de datos- proporciona herramientas para vincularlos tales como identificadores persistentes y citas de datos.

- En la medida de lo posible, los datos de investigación deben estar disponibles para otros usuarios, quienes podrán acceder, explotar, reproducir y difundir sin costo. Las licencias Creative Commons (CC BY o CC0) contribuyen en lograr este requisito (ver anexo 1).

3.1.3. ORD Pilot

Desarrollado en el contexto del programa Horizonte 2020 de la Unión Europea, el proyecto ORD Pilot¹⁸ tiene como objetivo mejorar y maximizar tanto el acceso como la reutilización de los datos de investigación generados por los proyectos nacidos al alero del mencionado programa, el más ambicioso a la fecha en lo que se refiere a innovación e investigación. Según sus bases, el proyecto piloto se caracteriza por dos pilares principales: desarrollar un Plan de Gestión de Datos (Data Management Plan o DMP) y proporcionar AA a los datos de investigación. Para participar, los interesados deben cumplir algunos requisitos: desarrollar un Plan de Gestión de Datos o DMP, el cual debe actualizarse según sea el caso; depositar los datos en un repositorio de datos de investigación; permitir que terceras personas puedan acceder a él y disponer de dichos datos para sus propios usos; y proporcionar información que permita relacionar o identificar los datos sin procesar para validar la información (OpenAIRE, s.f.)

ORD Pilot también surge de la European Association of Databases for Education and Training / Colaboration Data Infrastructure (EUDAT CDI)¹⁹, quien desarrolló B2SHARE²⁰. El uso de esta potente herramienta incluye un asistente de licencia incorporado que facilita la selección de una licencia adecuada para los datos de investigación. También sirve a los investigadores para almacenar, preservar y recuperar datos de variadas disciplinas.

3.1.4. B2SHARE

¿Qué características adicionales tiene B2SHARE sugerido por el ORD Pilot?

B2SHARE es un repositorio integrado con la infraestructura de datos colaborativos EUDAT y permite el depósito y registro en AA de datos de investigación. Este recurso permite a los investigadores o grupos de investigación definir la política de acceso de sus datos. También a los datos aquí

¹⁷ Zenodo es un repositorio de datos de investigación resultante de la colaboración entre OpenAIRE y el Conseil Européen pour la Recherche Nucléaire (CERN, en francés), de tipo generalista. Más datos en: <https://zenodo.org/>

¹⁸ ORD Pilot u Open Research Data Pilot (Piloto de Datos de Investigación Abiertos, en español). Sitio web: <https://www.openaire.eu/what-is-the-open-research-data-pilot>

¹⁹ También surge en el contexto de Horizonte 2020 para apoyar la Ciencia Abierta en la nube. Más datos en: <https://eudat.eu/>

²⁰ Sitio web: <https://b2share.eudat.eu/> y https://eudat.eu/services/userdoc/b2share#What_is_B2SHARE

almacenados se les asigna un identificador persistente, denominado ePersistent Identifiers for eResearch (ePIC)²¹, extensiones de metadatos e interfaz de usuarios de acuerdo con comunidades disciplinares, interfaz de programación de aplicaciones (API REST) para la integración con sitios de la comunidad, aplicaciones y flujos de trabajo, integridad de datos garantizada por suma de verificación durante la ingesta de datos, módulo de estadísticas para descargas de set de datos y soporte de usuario, entre otras funciones. Utiliza el software de código abierto INVENIO²² el cual nació en Conseil Européen pour la Recherche Nucléaire (CERN)²³.

La Figura 3-3: Diagrama del repositorio de datos de investigación B2SHARE ilustra la arquitectura del software:

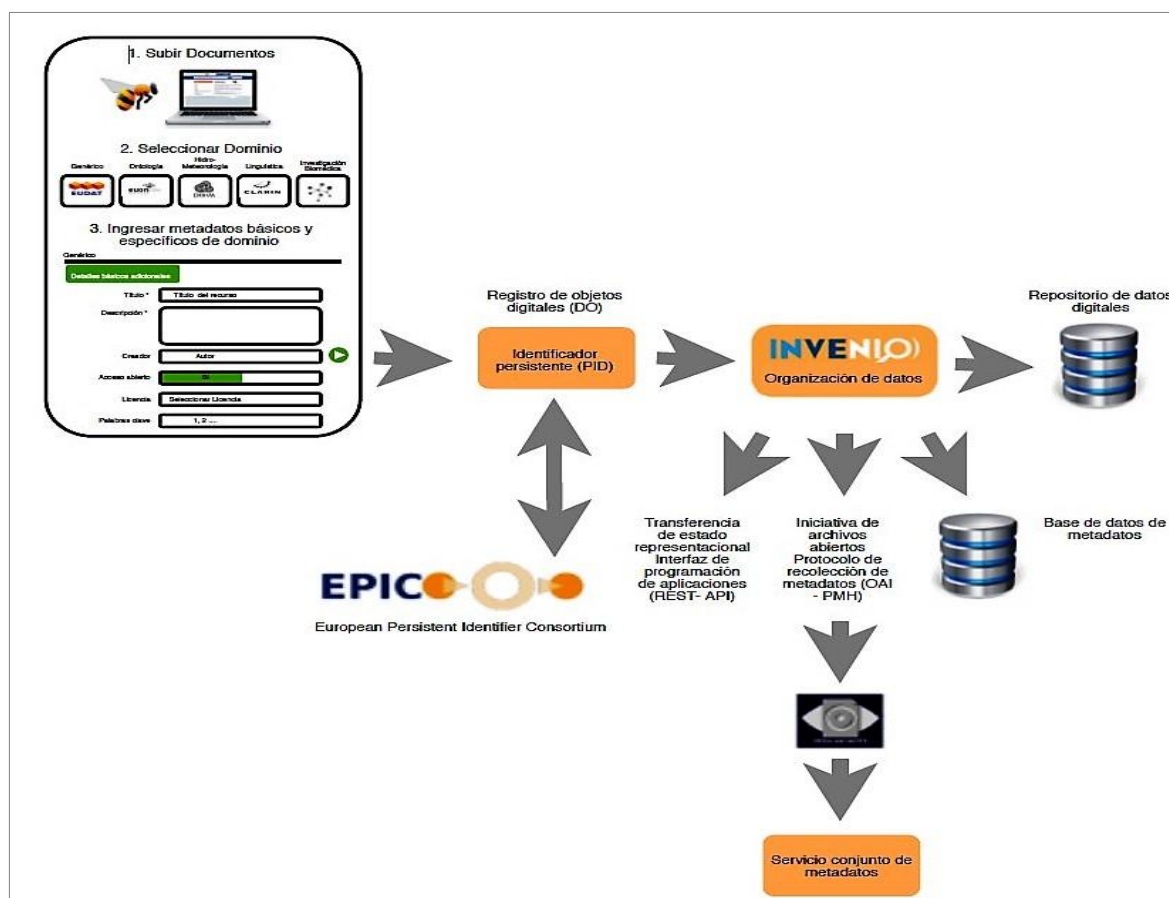


Figura 3-3: Diagrama del repositorio de datos de investigación B2SHARE

Fuente: <https://b2share.eudat.eu/>

²¹ ePIC es un consorcio fundado en 2009 e integrado por socios europeos con el fin de proporcionar servicios de identificadores persistentes para la CE. ePIC es un consorcio internacional y está abierto a socios de la comunidad de investigación de todo el mundo. Sitio web: <https://www.pidconsortium.net/>

²² El objetivo es impulsar la ciencia abierta. Sitio web: <https://invenio-software.org/>

²³ Conseil Européen pour la Recherche Nucléaire (CERN u Organización Europea para la Investigación Nuclear, en español). Fundada en 1954. Sitio web: <https://home.cern/>

En 2017, la CE generó también una guía, donde se explican los alcances respecto del AA para las publicaciones científicas revisadas por pares, así como para los datos de investigación, cuyas recomendaciones deben ser acogidas por los investigadores que hayan sido beneficiados en proyectos financiados por el programa Horizonte2020.

En este documento se destina un contenido llamado “Gestión de Datos Científicos y su Diseminación”, cuyas directrices emanadas por el consejo, en relación a los datos de investigación, exponen lo siguiente:

- El Consejo Europeo de Investigación (ERC)²⁴ apoya el principio básico de AA a los datos de investigación. Por lo tanto, “recomienda a todos los investigadores financiados por esta entidad, que sigan las mejores prácticas de conservación de archivos con los datos de investigación que hayan producido y utilizado durante el transcurso de su trabajo, y que estén dispuestos a compartirlos con otros investigadores, siempre que no estén sujetos a restricciones de derechos de autor, requisitos de confidencialidad o cláusulas contractuales”.
- El documento continúa: “Los requisitos legales para los proyectos que participan en el Piloto ORD se establecen en el Artículo 29.3 del convenio de subvención: el beneficiario debe depositar los datos en un repositorio de datos de investigación de manera que sea posible acceder, explotar, reproducir y compartir estos datos, libres de cargo para cualquier interesado. Para los proyectos del Programa de Trabajo ERC 2016, el Artículo 29.3 señala que el beneficiario solo será incluido en el Piloto ORD en tanto lo haya decidido así. A partir del programa de Proyectos de Trabajo 2017, el contenido del Artículo mencionado se incluirá por defecto en los Acuerdos de Subvención.”
- Respecto de los datos, prosigue el documento: “Los beneficiarios con recursos del ERC (Programa de Trabajo 2016) podrán optar, de manera individual y voluntaria, al Piloto de Horizonte 2020 sobre datos abiertos de investigación, con el fin de facilitar el acceso, la reutilización y la conservación de los datos de investigación generados durante su trabajo de investigación. Los que escojan esta opción, deben chequear cuidadosamente las obligaciones adicionales aplicadas a los proyectos que se adhieran al mencionado piloto, tal como se describe en el Artículo 29.3 del Modelo ERC antes señalado. A partir del programa de trabajo 2017, el Piloto para los datos de investigación comenzará a ampliarse hasta abarcar todas las áreas temáticas de Horizonte 2020. El AA se convierte así en un valor predeterminado por defecto para el conjunto de los datos de investigación generados. Los beneficiarios pueden optar en cualquier momento o etapa, liberándose de cualquier obligación relativa al AA a los datos de investigación producidos” (ERC. European Research Council, 2017).

²⁴ European Research Council (ERC, en inglés). Sitio web: <https://erc.europa.eu/>

Europa presenta y está trabajando en un ambiente dirigido hacia la colaboración entre los países generadores de una creciente investigación científica. En este ambiente irrumpe también en el Plan S²⁵, el cual representa un rol para la CA a nivel mundial, donde las editoriales más grandes no están dispuestas a ceder sobre sus intereses. Por esta razón, la Coalición S (de donde emana el Plan S) – con 24 integrantes a la fecha, entre los cuales se encuentran la Organización Mundial de la Salud, la Comisión Europea (CE), el Consejo Europeo de Investigación, la United Kingdom Research and Innovation (UKRI), la Fundación Bill y Melinda Gates, el Programa para la Investigación y Entrenamiento en Enfermedades Tropicales (TDR), haciendo notar que a éste último se encuentran vinculados la UNICEF y el Banco Mundial, y el Programa de las Naciones Unidas para el Desarrollo (UNDP)- requiere seguir sumando adherentes a la iniciativa²⁶ (López, 2020), la cual se circunscribe a diez principios, dentro de los cuales se destacan los siguientes: los autores mantendrían el copyright de sus publicaciones (mientras se encuentren en la modalidad de una licencia abierta tal como CC); proporcionar incentivos para la creación de revistas de AA y plataformas; y los costos de publicación deberían ser asumidos por los financiadores, institutos de investigación o universidades, en vez de los investigadores individuales, costos que se sugiere sean estandarizados y limitados en un entorno caracterizado por el alineamiento de políticas y de estrategias de las instituciones dispuestas a asumir dichos costos.

3.2. Presentación Ciencia Abierta y Acceso Abierto en Estados Unidos

- **2003: La Declaración de Bethesda sobre Publicación de Acceso Abierto**²⁷, suscrita en 2003 en Maryland, es un destacado hito mencionado en el levantamiento de la investigación documental como parte del AA en Estados Unidos.
- **2009: Las Academias Nacionales de Ciencias** hicieron en 2009 un llamado para disponer de mejores estándares que permitieran el intercambio para apoyar la reproducibilidad a través del intercambio ético de datos junto con los resultados de investigación publicados (Baker, 2017).
- **2013: Iniciativa de Gobierno Abierto** (Open Government Initiative, OGI), firmada el 9 de mayo de 2013 por el ex presidente Barack Obama, convirtió los datos abiertos y legibles por máquina en el nuevo valor predeterminado para la información del Gobierno. “Mi administración está comprometida a crear un nivel sin precedentes de apertura en el Gobierno. Trabajaremos juntos para asegurar la confianza pública y estabilidad de la transparencia del sistema, participación pública y colaboración. La apertura fortalecerá nuestra democracia y promoverá eficiencia y eficacia en el Gobierno” (Obama, 2013).

En esta senda han existido posteriores iniciativas para los datos abiertos en variados sectores, tales como salud, energía, clima, educación, finanzas, seguridad pública, y desarrollo global²⁸. En este

²⁵ Plan S. Promueve el Acceso Abierto completo e inmediato como una realidad. Sitio web: <https://www.coalition-s.org/>

²⁶ Consultado 290620.

²⁷ Declaración de Bethesda sobre publicación de Acceso Abierto. Sitio web: https://ictlogy.net/articles/bethesda_es.html

²⁸ Open Data Highlights. Sitio web:

contexto, la Casa Blanca también lanzó el Proyecto de Datos Abiertos (Project Open Data)²⁹ un repositorio que tiene por objetivo compartir las mejores prácticas, ejemplos de casos y softwares de código abierto. Para ser consecuentes con la iniciativa, todo se publicaría en GitHub³⁰.

La administración de Obama dio una señal en este tema, recomendando que la investigación pública financiada por los contribuyentes estuviese a disposición de todas las personas, después de un año de ser publicada.

Este mandato para compartir y proporcionar acceso éticamente abierto a los datos pretendía impulsar un cambio en las prácticas de datos generados por la investigación en diferentes áreas, incluyendo aquí a los repositorios. Se tenía la noción de que el ecosistema de datos requería de un trabajo colaborativo desde diferentes instancias para integrar los ámbitos sociales, técnicos, organizativos y políticos (Baker, 2017).

- **2015: El Gobierno de Estados Unidos entrega aportes a las Agencias Federales**³¹. En este contexto, la Oficina de Política de Ciencia y Tecnología³² (OSTP) incentivó a las agencias enfocadas en temas de ciencias para proporcionar acceso a información y datos científicos obtenidos con financiamiento público, lo cual sería un nuevo incentivo para la innovación, la economía y el bienestar social.
- **2015-2017: Iniciativa “Open Science Through Open Data (US0072)”**³³. Constituida con representantes gubernamentales y de la sociedad civil, tiene como propósito la promoción de iniciativas para este tema, propiciando su desarrollo en ambientes de altos estándares que aseguran, de forma complementaria, la sostenibilidad en el largo plazo.
- **2018-2020: Oficina de Política de Ciencia y Tecnología (OSTP)**. Dentro de su estrategia de mejoras de acceso público a la investigación financiada por contribuyentes (tema inspirado por el memorando de 2013), esta repartición ha sostenido numerosas reuniones con editoriales –incluyendo aquí a la Asociación de Editores Estadounidenses-, universidades, investigadores entre otros, para abordar este tema, el cual es fundamental y previo a la generación de una política emanada desde la Casa Blanca. Los avances no han estado exentos de polémicas, ya que las editoriales no comparten ceder sus publicaciones sin costo. Esta controversia y postura es respaldada por más de ciento veinticinco grupos editoriales, incluida la Cámara de Comercio de los Estados Unidos. Esto se traduce en que el tema del Acceso Público (denominado así en este país) está en una actual controversia, enfrentando a los principales editores del país y a las sociedades académicas más pequeñas, contra los

<https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2013opendata.pdf>

²⁹ Sitio web: <https://obamawhitehouse.archives.gov/blog/2013/05/16/introducing-project-open-data>

³⁰ Plataforma de desarrollo de amplio alcance, que incluye desde el código abierto hasta las necesidades de las instituciones, permitiendo alojar, administrar proyectos y crear software. Sitio web: <https://github.com/>

³¹ Las Agencias Federales son independientes del Gobierno Federal de los Estados Unidos. En la actualidad, totalizan 40 oficinas, cubriendo diferentes áreas.

³² La Oficina de Política de Ciencia y Tecnología (Office of Science and Technology Policy, OSTP) es una repartición del Gobierno de los Estados Unidos, que integra la Oficina Ejecutiva del Presidente (EOP), establecida por el Congreso de Estados Unidos el 11 de mayo de 1976.

³³ Sitio web: <https://www.opengovpartnership.org/es/members/united-states/commitments/US0072/>

defensores del Acceso Público con fondos provenientes de todos los contribuyentes de la nación.

En la sección Notices, del documento Federal Register (Vol. 85, No. 33 / Wednesday, February 19, 2020), se publicó la siguiente información de la OSTP: “OFFICE OF SCIENCE AND TECHNOLOGY POLICY Request for Information: Public Access to Peer-Reviewed Scholarly Publications, Data and Code Resulting from Federally Funded Research”³⁴.

3.2.1. Agencias Federales – Institutos Nacionales de Salud (NIH) como ejemplo

Como las Agencias Federales tienen un rol destacado para promover el AA de los resultados alcanzados con financiamiento de los contribuyentes, entregando en acceso público esta información, a continuación se presenta el caso de los Institutos Nacionales de Salud (NIH, en inglés), quienes han incursionado y disponen de buenas experiencias y prácticas en la implementación de infraestructura tecnológica.

- **2012-2020: La Oficina de Estrategia de Ciencia de Datos (ODSS, en inglés)**³⁵, dependiente de los NIH³⁶, éstos últimos dependientes a su vez del Departamento de Salud y Servicios Humanos de los Estados Unidos³⁷, lidera la implementación del Plan Estratégico de NIH³⁸ para Ciencia de Datos, el cual fue lanzado en junio de 2018 con el propósito de almacenar, administrar, estandarizar y publicar la gran cantidad de datos producidos por la investigación biomédica (NIH. National Institutes of Health, s.f.).

Este plan estratégico proporciona una hoja de ruta, la cual tiene por objetivo renovar el ecosistema de ciencia de datos biomédicos, al cual aportan veintisiete institutos y centros financiados por los NIH. Por otra parte, este plan pone énfasis en que los datos accesibles, bien organizados, seguros y operados de manera eficiente, considerados fundamentales para el avance de la investigación, lo cual redundará en un rápido progreso de la ciencia.

Un fundamento esencial del plan es que todos los datos de investigación biomédica deben adherirse a los Principios FAIR³⁹, lo cual significa que los datos deben ser encontrables (o localizables), accesibles, interoperables y reutilizables.

³⁴ Federal Register 2020. PDF en sitio web: <https://www.govinfo.gov/content/pkg/FR-2020-02-19/pdf/2020-03189.pdf>

³⁵ Office of Data Science Strategy (ODSS). Sitio web: <https://datascience.nih.gov/>

³⁶ National Institutes of Health (NIH). Sitio web: <https://datascience.nih.gov/>

³⁷ Departamento de Salud y Servicios Humanos. Sitio web: <https://www.hhs.gov/>

³⁸ Plan Estratégico NIH PDF en sitio web: https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf

³⁹ FAIR o Principios rectores para la gestión y administración de datos científicos. Sitio web: <https://www.go-fair.org/fair-principles/>

Se destaca que esta Oficina promueve, además, colaboraciones innovadoras entre la ciencia de datos biomédicos con otras áreas, tales como matemáticas, estadísticas, informática, ingeniería, física y biología para crear modelos sinérgicos de descubrimientos basados en datos⁴⁰.

Para lograr esto, los NIH disponen de una robusta infraestructura de datos que debe garantizar la optimización del almacenamiento y de la seguridad de los datos, propiciando también conexiones de sistema de datos que forman parte del ecosistema sustentado por los NIH (NIH. National Institutes of Health, s.f.).

Con el fin de incorporar los principios -cuya finalidad es disponer en acceso público los resultados obtenidos con financiamiento gubernamental- la organización también declara un capítulo para la **Participación ciudadana**, donde ha tomado fuerza la **Ciencia Ciudadana**⁴¹, cuya definición corresponde a la participación del público en general en actividades de investigación científica en las que los ciudadanos contribuyen activamente, ya sea con su esfuerzo intelectual o con el conocimiento de su entorno o aportando sus propias herramientas y recursos (Natusfera, 2017).

La otra vía dispuesta para la participación ciudadana es **Code-a-thons**⁴², cuyo objetivo es organizar encuentros entre variados investigadores y programadores quienes pueden trabajar juntos y crear herramientas y canales de tutorías que tienen por misión resolver problemas relevantes para la comunidad biomédica. De este modo se potencia la colaboración, reafirmando el interés de que los resultados alcanzados sean además compartidos a todos los estamentos de la ciudadanía.

En este ámbito, esta organización cuenta con “Políticas de intercambio de NIH y orientación relacionada sobre recursos de investigación financiados por el Instituto”⁴³. En síntesis, esta política releva que los resultados y logros de las actividades con fondos del NIH deben estar a disposición del público que lo requiera (NIH. National Institutes of Health, s.f.).

3.3. Presentación Ciencia Abierta y Acceso Abierto en América Latina

Un estudio cualitativo exploratorio se centró en analizar el estado de situación de la CA en América Latina en los últimos diez años, respecto de la implementación de políticas públicas en esta materia. Sólo siete países demostraron iniciativas: Argentina, Brasil, Chile, Colombia, México, Perú y Uruguay. Asimismo, este estudio dejó en evidencia que gran parte de estas políticas se enfocan en la

⁴⁰ Sitio web: <https://datascience.nih.gov/community-engagement>

⁴¹ El Biomedical Citizen Science Hub, cuya misión es crear un espacio de colaboración en línea para los recursos, proyectos, referencias, métodos y comunidades biomédicas de ciencia ciudadana, biomédicos en crecimiento y prácticamente dispersos para que los interesados los descubran y participen. Más datos en: <https://citscibio.org/>

⁴² Code-a-thons es un sitio de la Biblioteca Nacional de Medicina de los Estados Unidos (NLM) para reunir a personas con diversos antecedentes con el fin de crear herramientas para el análisis avanzado de datos biomédicos. Sitio web: <https://ncbi-codeathons.github.io/>

⁴³ Sitio web: <https://grants.nih.gov/policy/nihgps/index.htm>

promoción de infraestructuras para el AA, en particular a publicaciones⁴⁴ (De Filippo & y D’Onofrio, 2019).

Dentro de las iniciativas que son mencionadas con mayor frecuencia en la literatura por estar en sintonía con la CA y el AA en América Latina figuran: Latindex, Scielo, AmeliCa, Redalyc y LA Referencia. Sin embargo, hasta la fecha, todas ellas se encuentran orientadas más bien hacia las publicaciones.

3.3.1. Latindex



Figura 3-4: Portal latindex

Fuente: <https://www.latindex.org/latindex/inicio>

Latindex es un Sistema Regional de Información en Línea para Revistas Científicas de América Latina, El Caribe, España y Portugal. Nació en 1995 en la Universidad Nacional Autónoma de México (UNAM), institución que en 1997 se convirtió en una red de cooperación regional. Es un portal de revistas, el cual cuenta con 23 instituciones asociadas del continente, una institución de Europa y una de Asia (UNAM. Universidad Nacional Autónoma de México, 2018).

A la fecha, en su portal no manifiesta información respecto de los datos de investigación.

⁴⁴ Artículo completo en: <https://www.raco.cat/index.php/Hipertext/article/view/360106> Consultado 230620.

3.3.2. SciELO



Figura 3-5: Portal SciELO

Fuente: <https://scielo.org/es/>

La Red SciELO nació en Brasil en 1998. Actualmente, está integrada por 16 países preferentemente del continente. Dispone de una declaración de AA para los contenidos de revistas científicas, las cuales se organizan en colecciones nacionales y temáticas. Es responsabilidad de cada país gestionar la selección y mantenimiento de las revistas que integran la colección. Las revistas publican en SciELO artículos de investigación, de revisión, comunicaciones vinculadas con la investigación, estudios de caso; sin embargo, hasta la fecha, no se hace mención a los datos de investigación que pudiesen ser constitutivos de las publicaciones.

Creative Commons (CC) es la organización que formaliza en el portal el AA. Así, también, desde 2015 la licencia CC-BY⁴⁵ fue adoptada por los Criterios SciELO de indexación como el estándar de asignación para AA. Las publicaciones indexadas antes del año mencionado tuvieron la posibilidad de mantener su licencia original; sin embargo, se les solicitó adoptar CC-BY con el propósito de hacer uso de todos los servicios de interoperabilidad y diseminación de la red.

Esta licencia CC-BY 4.0 permite que otros distribuyan, remienden, adapten y creen a partir de su trabajo, incluso con fines comerciales, siempre que le asignen el debido crédito por la creación original. Es la licencia más flexible de todas las licencias disponibles. Se recomienda para maximizar la difusión y el uso de los materiales licenciados (SciELO. Scientific Electronic Library Online, s.f.).

⁴⁵ Se trata de un grupo de licencias de derecho de autor abiertas y/o libres. Sitio web: <https://creativecommons.org/licenses/>

3.3.3. AmeliCA

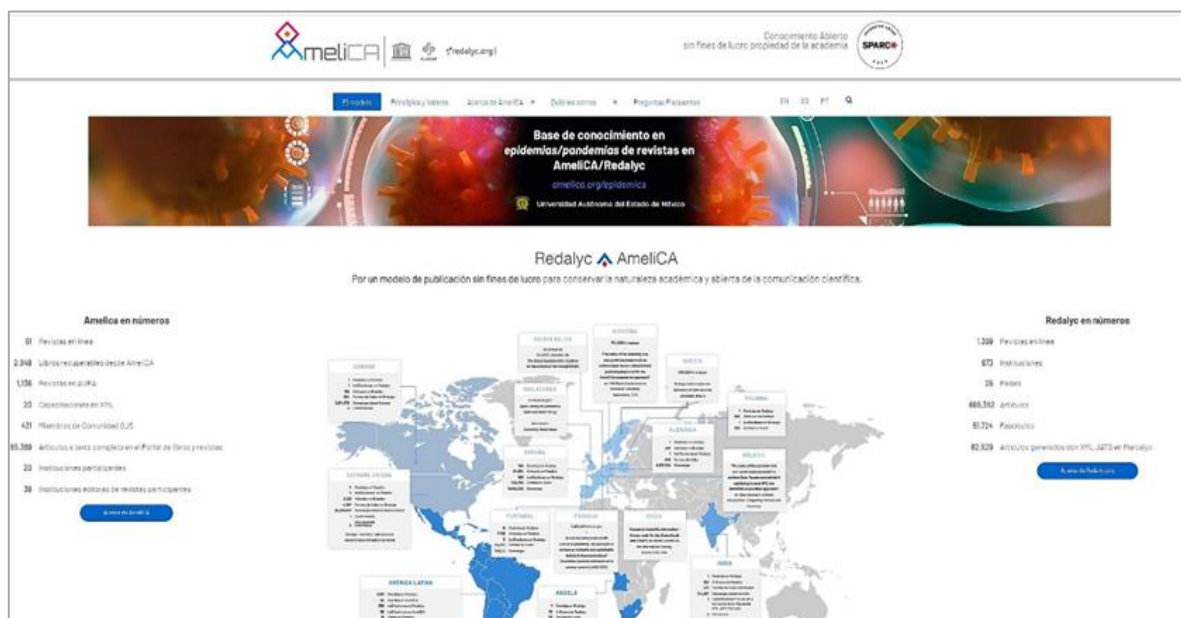


Figura 3-6: Portal AmeliCA

Fuente: <http://amelica.org/>

AmeliCA se fundó en 2018 como una infraestructura de comunicación para la publicación académica y la CA, sostenida de forma cooperativa y centrada en un modelo de publicación sin fines de lucro, con el propósito de conservar la naturaleza académica y abierta de la comunicación científica. La iniciativa es liderada por la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), el Consejo Latinoamericano de Ciencias Sociales (CLACSO), Red de Revistas Científicas de América Latina y el Caribe, España y Portugal (Redalyc) e impulsada -desde su lanzamiento- por la Universidad Autónoma del Estado de México (UAEM) y la Universidad Nacional de La Plata (UNLP) (AmeliCA, s.f.).

Declara en su web que CA es un eje para promover el proceso de comunicación científica en cada una de sus fases de manera abierta y colaborativa, tanto en lo referente a los procesos e instrumentos de trabajo, como a los resultados intermedios y finales del proceso de investigación. En ello, se estimula la colaboración con otros actores generadores de conocimiento, más allá del ámbito académico, como son la ciudadanía y gobiernos. En 2019 recibe el premio Scholarly Publishing and Academic Resources Coalition (SPARC)⁴⁶ a la innovación.

⁴⁶ SPARC es una Iniciativa que reconoce a un individuo, institución o grupo por ejemplificar los principios de SPARC al trabajar para desafiar el statu quo en la comunicación académica en beneficio de investigadores, bibliotecas, universidades y el público. Sitio web: <https://sparcopen.org/our-work/innovator/>

3.3.4. Redalyc



Figura 3-7: Portal Redalyc

Fuente: <https://www.redalyc.org/>

La Red de Revistas Científicas de América Latina y el Caribe, España y Portugal (Redalyc) -nacida en 2002 en la UAEM- continúa sumando numerosos colaboradores desde universidades hasta editoriales internacionales, perfilándose como un proyecto académico para la difusión en AA de la actividad científica editorial de todo el mundo, bajo un modelo liderado por la academia y no lucrativo. Actualmente, este portal permite indizar, visibilizar, compartir y preservar la producción editorial. También proporciona indicadores a las publicaciones aquí alojadas.

Recientemente, en 2020, ante la amenaza del sistema editorial latinoamericano por el incentivo, promoción y apoyo a las revistas -en función de indicadores bibliométricos y ante el posicionamiento y apoyo en favor del modelo Article Processing Charge (APC)⁴⁷ que supone el Plan S-, Redalyc adhirió a modo de respuesta a un sistema de publicación abierto, propiedad de la academia, no lucrativo, sin fines de lucro y sostenible para conservar la naturaleza académica y abierta de la comunicación científica de cualquier región. Las publicaciones que eligen este modelo pueden optar a “calidad editorial y científica (revisión por pares), tecnología de publicación digital XML-JATS, una política de AA sin costos por publicación o procesamiento y la visión de superar la actual evaluación de la ciencia basada en métricas como el Factor de Impacto, e impulsando la inclusión de la ciencia local y la diversidad lingüística por el bien común” (UAEM. Universidad Autónoma del Estado de México, 2020).

⁴⁷ Article Processing Charge (APC) son cargos por procesamiento de artículos. Sitio web: <https://blog.scielo.org/es/tag/article-processing-charge-apc/>

3.3.5. LA Referencia



Figura 3-8: Portal LA Referencia

Fuente: <http://www.lareferencia.info/es/>

La Red Federada de Repositorios Institucionales de Publicaciones Científicas (LA Referencia)⁴⁸ tiene como misión ofrecer en AA la producción científica de América Latina financiada con fondos públicos, mediante la cooperación y articulación de una red federada de repositorios institucionales, basada en acuerdos regionales y estrategias nacionales de AA. Dentro de su red de apoyo están OpenAIRE y COAR⁴⁹. OpenAIRE proporciona a LA Referencia una infraestructura tecnológica para la interoperabilidad e indexación estándar en los principales buscadores mundiales; por otra parte, la

⁴⁸ Red Federada de Repositorios Institucionales de Publicaciones Científicas (LA Referencia). Por medio de sus servicios, apoya las estrategias nacionales de Acceso Abierto en América Latina mediante una plataforma con estándares de interoperabilidad, compartiendo y dando visibilidad a la producción científica generada en las instituciones de educación superior y de investigación científica. Nace del [Acuerdo de Cooperación](#), firmado en Buenos Aires (2012), que refleja la voluntad política de ofrecer en acceso abierto la producción científica de América Latina como un bien público regional con énfasis en los resultados financiados con fondos públicos. Sitio web:

<http://www.lareferencia.info/es/institucional/quienes-somos> Consultado 220620.

⁴⁹ Confederation Open Access Repositories (COAR). Se creó en 2009 para conectar una red de repositorios europeos llamado DRIVER. Actualmente cuenta con 157 miembros y socios de todo el mundo que representan bibliotecas, universidades, instituciones de investigación, financiadores gubernamentales y otros. COAR reúne repositorios individuales y redes de repositorios para desarrollar capacidades, alinear políticas y prácticas, y actuar como una voz global para la comunidad de repositorios. Sitio web: <https://www.coar-repositories.org/about-coar/> Consultado 230620.

vinculación con COAR le permite formar parte de acuerdos específicos con iniciativas como Zenodo⁵⁰, del CERN⁵¹, para datos científicos.

En relación a políticas para la CA y los Datos Científicos en América Latina, LA Referencia impulsa el AA acogiendo como principio básico la convicción de que el conocimiento es un bien público. En lo que respecta a los datos de investigación, especialmente los financiados con fondos públicos sostiene:

- Reproducir y validar los resultados de mejor forma, generar nuevas investigaciones y conocimiento a partir de los datos científicos abiertos ya existentes.
- Dotar de mayor transparencia a los fondos aplicados a la investigación.
- Preservar la integridad de la investigación.
- Evitar la duplicación de esfuerzos (LA Referencia, s.f.).

Asimismo, en este documento se destaca a legislaciones nacionales de AA en Argentina y Perú, en 2013, y México, en 2014, las cuales incluyen a los datos científicos como parte de la política (LA Referencia, s.f.).

3.4. Presentación Ciencia Abierta y Acceso Abierto en Chile

La Agencia Nacional de Investigación y Desarrollo (ANID), nombre con el cual se conoce la continuación del Comisión Nacional de Investigación Científica y Tecnológica (CONICYT), se encontró en un proceso de consulta abierta hasta el 31 de julio de 2020. El propósito fue conocer las opiniones de la comunidad científica y de la ciudadanía, a través de una consulta participativa no vinculante, con el propósito de implementar a partir de 2021 la Política de Acceso Abierto a Información Científica y Datos de Investigación. A través de lo anterior se busca “asegurar la disponibilidad del conocimiento científico contenido en las publicaciones resultantes de la ejecución de los proyectos de investigación, financiados con recursos provenientes de la Agencia”⁵² (ANID. Agencia Nacional de Investigación y Desarrollo, 2020).

En este proceso de consulta, el documento propuesto declara que -para nuestro país- los ejes del ecosistema digital de información científica (liderados por la ANID) corresponden a los presentados en la Fig. 9⁵³ (ANID. Agencia Nacional de Investigación y Desarrollo, 2020).

⁵⁰ Zenodo fue creado el 2013 y se define como un repositorio de Acceso Abierto de uso general desarrollado bajo el programa europeo OpenAIRE y operado por el CERN. Permite a los investigadores depositar conjuntos de datos, softwares de investigación e informes relacionados con la investigación. Para cada envío se acuña un Identificador de Objeto Digital (DOI), lo que hace que los elementos almacenados sean fácilmente citables. Sitio web: <https://en.wikipedia.org/wiki/Zenodo> Consultado 230620.

⁵¹ Consejo Europeo para la Investigación Nuclear (Conseil Européen pour la Recherche Nucléaire, CERN). Existe desde 1954 y se define como una organización intergubernamental. Aloja a Zenodo y, con esto, reafirma ser reconocida por su trabajo pionero en Acceso Abierto. Sitio web: <https://about.zenodo.org/infrastructure/> Consultado 230620.

⁵² Consultado 250620.

⁵³ Consultado 250620.



Figura 3-9: Ejes en Chile sobre los que se sustenta el Acceso Abierto

Fuente: adaptación ANID

La consulta propone que los actuales ejes sean articulados bajo una sola política, acercándose así a la mirada existente en Europa respecto de esta materia, continente donde el AA es un relevante bastión para las sociedades, siendo una instancia para fortalecer las economías, transparentar y dar visibilidad a la producción científica, permitir el acceso y compartir los resultados íntegros de la investigación.

Además, es importante articular dicha política porque, en los ejes del ecosistema digital de investigación, no se han abordado de modo explícito los datos científicos, tema que tiene gran impacto. Para dar sustento a esta mirada, Remedios Melero y Ernest Abadal⁵⁴ (Melero, 2018) señalan que los datos de investigación (o datos científicos) despiertan gran interés por su potencial aprovechamiento y reutilización, no sólo por parte del sector de la I+D⁵⁵, sino también por parte de otros agentes como la industria o las empresas de servicios que pueden utilizarlos para la innovación de sus productos o para la creación de nuevos empleos.

Desde el punto de vista de la investigación, compartir y acceder a los datos generados durante la actividad investigadora supone múltiples beneficios a nivel institucional e individual. En las instituciones proporciona transparencia en los procesos de obtención o generación de datos. Entre los investigadores, promueve la colaboración a nivel de grupos afines e interdisciplinarios, evitando así duplicaciones. Para la sociedad, los beneficios de hacer disponibles los datos mejoran la confianza en el sistema científico y suponen un ejercicio de transparencia, rendición de cuentas y de responsabilidad en el uso de la inversión en ciencia. Consideran que la colaboración es fundamental para abordar la compleja infraestructura tecnológica que acompaña a la ciencia.

⁵⁴ Consultado 270620.

⁵⁵ Y+D Símbolo que alude a Investigación y Desarrollo, proceso que tiene por objetivo el desarrollo de tecnologías con el fin de generar nuevos productos, materiales o procesos, donde resultan esenciales la creatividad y novedad. Sitio web: <https://economipedia.com/definiciones/investigacion-desarrollo-id.html>

Retomando la propuesta de esta política de AA para Chile, la ANID declara -en parte de su objetivo general- incluir los datos de investigación, hecho que confirma en uno de sus objetivos específicos: “Facilitar la reutilización de la información y los datos de investigación mediante estándares de interoperabilidad para generar nuevo conocimiento”⁵⁶ (ANID. Agencia Nacional de Investigación y Desarrollo, 2020).

Por otra parte, la propuesta alude a que se abordará en dos fases la Política de AA: una fase para las publicaciones y otra fase (de interés para abordar en esta publicación) enfocada en los datos, basándose en los Principios FAIR⁵⁷.

La propuesta en cuanto a la gestión e implementación dispone, además, que la Subdirección de Redes, Estrategia y Conocimiento (REC) de la ANID deberá diseñar e implementar lo que sea necesario para que se cumpla la política. De este modo, se orientará a los investigadores y es importante que tanto bibliotecas como instituciones que están por implementar o mejorar sus repositorios de datos estén en conocimiento de lo que esto implicará. A modo de referencia:

Por un lado, estar atentos a los estándares de interoperabilidad con que opera el Repositorio de Producción Científica de ANID (si su institución está vinculada), ya que éstos -además de ser difundidos- serán actualizados. Por otro lado, estar atentos al nuevo establecimiento de estándares de interoperabilidad para datos, si está en el plan de su institución la implementación de un repositorio de datos.

3.5. Softwares para repositorios de datos de investigación

A continuación, y tomando como referencia la bibliografía especializada disponible, más las experiencias recogidas por parte de los usuarios (recogidas a través de la metodología que se expone en otro capítulo en detalle), documentación y entrevistas con expertos en la materia, describo los principales softwares para repositorios de datos de investigación usados actualmente en diversas instituciones en el mundo. En cada caso, proporciono una breve reseña histórica sobre su origen, los objetivos por los cuales fue concebido, las características técnicas y algunos ejemplos significativos de usuarios que lo han implementado. Los softwares fueron dispuestos según criterios estadísticos, comenzando por el más requerido en el mercado.

3.5.1. DSPACE

⁵⁶ Consultado 250620.

⁵⁷ Principios FAIR. Enfatizan la capacidad de acción de la máquina (es decir, la capacidad de los sistemas computacionales para encontrar, acceder, interoperar y reutilizar datos con ninguna o mínima intervención humana). Sitio web: <https://www.go-fair.org/fair-principles/> Consultado 270620.

	DSPACE https://duraspace.org/dspace/	
Origen	2002: Desarrollado por el Instituto de Tecnología de Massachusetts (MIT) en alianza con Hewlett-Packard (HP). Es liberado bajo una licencia BSD ⁵⁸ , lo cual facilita personalizar o extender softwares según se necesite.	
Objetivo	-Centralizar, normalizar, almacenar, diseminar y preservar la producción científica y académica en variados formatos generadas en las instituciones.	
Experiencias	Edinburgh DataShare, Apolo (Cambridge), Canadian Epigenetics, Environment and Health Research Consortium Network (CEEHRC), edoc-Server Open-Access-Publikationsserver der Humboldt-Universität, DRYAD, Nottingham Research Data Management Repository. En Chile, veinte instituciones entre gubernamentales y académicas disponen de este software, pero sólo una está incursionando con una colección para datos de investigación (Pontificia Universidad Católica de Chile).	
Características	Disponible gratuitamente como software de código abierto. Personalizable para: Interfaz de usuario dispone de una interfaz tradicional y también dispone de interfaces temáticas listas para habilitar. Metadatos Dublin Core es el formato de metadatos predeterminado; no obstante, se puede agregar o modificar cualquier campo. Configurable para la exploración por autor, título, fecha, etc., y para la búsqueda se puede seleccionar cualquier campo de metadatos que requiera incorporar en esta interfaz. Mecanismos de autenticación local para la institución LDAP (y LDAP jerárquico), Shibboleth, X.509, basado en IP e incorpora autenticación interna, o puede configurarse para usar múltiples métodos de autenticación a la vez. Compatibilidad de estándares, protocolos estándares de acceso, ingesta y exportación son OAI-PMH, OAI-ORE, SWORD, WebDAV, OpenSearch, OpenURL, RSS, ATOM. Base de datos configurable PostgreSQL u Oracle para administrar los metadatos.	

⁵⁸ BSD (Berkeley Software Distribution). Cualquier organización puede usar, modificar e incluso integrar el código en su aplicación comercial sin pagar ninguna tarifa de licencia. Sitio web: <https://duraspace.org/dspace/about/features/>

	Idioma predeterminado del software: disponible en más de veinte idiomas y se puede configurar para varios idiomas. La estructura permite organizar en comunidades y segmentar por tipo de colecciones. Ej. Colección de datos de investigación. Sistema operativo configurable para Linux, Mac OSX o Windows. Información sobre requisitos previos de software en: https://wiki.lyrasis.org/display/DSDOC/All+Documentation Gestionar y preservar todo tipo de contenido, acepta y reconoce muchos formatos, archivos PDF, Word, JPEG, MPEG, TIFF y puede administrar archivos de cualquier formato. Proporciona además un registro de formato de archivo simple, lo cual permite registrar cualquier formato no reconocido. Estadísticas recopiladas en la aplicación SOLR (sólo tabular). No hay límite para los archivos que se suben. Compatible con Google Analytics, para lo cual la API de GA recupera desde Google y genera informes. Identificador persistente HDL Handle.Net de Corporation for National Research Initiatives (CNRI)⁵⁹. Permite recibir alertas a través de protocolo RSS. Además, se usa para publicaciones y conjuntos de estándares (DSpace, 2020).	
--	--	--

3.5.2. Dataverse

	Dataverse https://dataverse.org/
Origen	1999-2006: Comienza el proyecto Dataverse desarrollado por el equipo Dataverse en el Instituto de Ciencias Sociales Cuantitativas (IQSS) de la Universidad de Harvard, el cual también contó con la colaboración de la Biblioteca de la mencionada casa de estudios superiores. Al tratarse de un

⁵⁹ Corporation for National Research Initiatives, CNRI (Corporación para las Iniciativas Nacionales de Investigación, en español). Es una organización sin fines de lucro que nace en 1986 para emprender, fomentar y promover la investigación en interés público. Sus actividades se centran en el desarrollo estratégico de tecnologías de información basadas en redes, proporcionando liderazgo y financiación para la investigación y desarrollo de infraestructura de información. Administrador mundial para identificadores como Handle.Net Registry, que permite asignar, administrar y resolver identificadores persistentes para objetos digitales y otros recursos de Internet. Sitio web: <http://www.cnri.reston.va.us/> Ver también: <https://www.handle.net/>

	proyecto, recibe también la colaboración internacional de muchas instituciones.
Objetivos	-Aumentar la adopción (usuarios, versículos de datos, conjuntos de datos, instalaciones, diarios). -Finalizar las funciones de migración de Dataverse 4. -Desarrollar la capacidad para manejar datos confidenciales, de gran escala y de transmisión de nivel 3. -Expandir las características de datos y metadatos para disciplinas existentes y nuevas. -Ampliar las funciones de archivo y preservación. -Aumentar las contribuciones de la comunidad de desarrollo de código abierto. -Mejorar UX y UI. -Continuar aumentando la calidad del software (Dataverse Project, 2020).
Experiencias	QDR: Repositorio de datos cualitativos (Estudios Unidos), heiData, Universidad de Ottawa, Archivo de Datos de Australia, Harvard Dataverse, Dataverse de la Universidad de Queen, Dataverse de las Bibliotecas de la Universidad de Alberta, Portal de datos abiertos de la Pontificia Universidad Católica del Perú, DataverseNO (Noruega) y Repositorio de Datos de Investigación de la Universidad de Chile (Chile).
Características	Aplicación web de código abierto para compartir, preservar, citar, explorar y analizar datos de investigación. El software aloja múltiples archivos llamados Dataverses, cada uno de los cuales contiene conjuntos de datos. A su vez, cada uno de dichos conjuntos contiene metadatos descriptivos y archivos de datos, incluyendo éstos la documentación y el código que acompaña a los datos. Personalizar e integrar en sitios web los datos alojados a través de widget⁶⁰, los cuales se pueden incrustar facilitando la localización de los datos de un investigador, esto repercute en una mayor visibilidad y, en la posibilidad de crear una cita estándar con siete componentes⁶¹ del conjunto de datos de un investigador brindando el reconocimiento y crédito de los autores. La cita de datos también está desarrollada para conjuntos de datos y archivos EndNote XML, formato RIS o formato BibTeX Exención CC0 para conjuntos de datos, a todos los conjuntos de datos agregados en el software se les otorga la dedicación de dominio público CC0 de forma predeterminada (4.0 y en adelante).

⁶⁰ Los widgets son micro aplicaciones interactivas para facilitar el acceso a ciertas funciones de otras aplicaciones.

⁶¹ Nombre(s) del autor, año (fecha publicada en Dataverse), título, Identificador Persistente Global (DOI) o Handle, editor (repositorio que publicó el conjunto de datos), número de versión, huella digital universal (UNF) para datos tabulares. Sitio web: <https://dataverse.org/best-practices/data-citation>

	El software también puede permitir articular un plan de gestión de datos⁶², para esto utiliza la herramienta DMPtool⁶³ y con esto facilitar a los investigadores cumplir con los requisitos de la agencia de financiación. El software incorpora también soporte para los Principios FAIR⁶⁴ (Dataverse Project, 2020). Estándares OAI-PMH (cosecha). Permiten recopilar y entregar metadatos desde y hacia otros sistemas, utilizando formatos estandarizados tales como Dublin Core, Data Document Initiative (DDI), OpenAIRE, entre otros. Estadísticas incorpora Make Data Count⁶⁵ herramienta para recopilar y estandarizar métricas sobre el uso de los datos, vistas, descargas y citas de archivos y conjuntos de datos. Descarga de archivos en formato R⁶⁶ y TSV⁶⁷ así como integrado a Dropbox.
--	--

3.5.3. CKAN

 ckan	CKAN (Comprehensive Knowledge Archive Network) https://ckan.org/
Origen	2016: Se desarrolla por la Fundación del Conocimiento Abierto (OKF) ⁶⁸ .
Objetivo	-Proporcionar una plataforma que permita catalogar, almacenar y acceder a conjuntos de datos con una interfaz completa (OKF. Open Knowledge Foundation, 2020) .
Experiencias	Datos de investigación de Australia, Portal de datos CSIRO (Australia), Infraestructura Europea de Datos (EUDAT), B2FIND (UE), Repositorio de Datos Científicos (USA), OpenData Dinamarca y muchos gobiernos que han hecho de este software su elección para datos gubernamentales.
Características	Software de código abierto, gratuito y que se puede adaptar el código libremente si es necesario. Dispone de una robusta API de tipo RPC, lo cual es útil para trabajar con una vista estrecha de los datos.

⁶² Plan de Gestión de Datos (Data Management Plan, DMP). Es un documento formal que detalla cómo se gestionarán los datos durante la investigación y una vez finalizado el proyecto. Es requerido por todas las agencias de financiación de proyectos. Permite asegurar la integridad y reproducibilidad de los datos, difundir y preservarlos, validar los resultados obtenidos e incrementar el impacto.

⁶³ Sitio web: <https://dmptool.org/>

⁶⁴ Sitio web: <https://scholar.harvard.edu/mercecosas/presentations/fair-guiding-principles-implementation-dataverse>

⁶⁵ Sitio web: <https://makedatacount.org/>

⁶⁶ R es un lenguaje de programación orientado al análisis de datos.

⁶⁷ Tab-Separated Values (TSV) sirve para para identificar los archivos basados en texto que contienen sólo eso: cadenas de valores que están separados por el carácter de tabulación horizontal (Código ASCII: 09H).

⁶⁸ Open Knowledge Foundation (OKF). Fundación sin fines de lucro, creada en Reino Unido en 2006. Sitio web: <https://okfn.org/>

	Permite a editores y usuarios de datos realizar la recolección de datos, la búsqueda por facetas y las interfaces de máquina para datos y metadatos, como también la federación, lo que le permite compartir sus datos públicos con otros repositorios desarrollados en CKAN. Interfaz de búsqueda y descubrimiento al estilo de “Google”, que permite también la creación de facetas por etiquetas y la navegación entre conjuntos de datos relacionados. Los usuarios pueden ver qué conjuntos de datos están disponibles, así como el formato y el tipo de licencia (vista instantánea en cuanto a si los datos están disponibles bajo una licencia abierta o no. Gracias a lo anterior, los usuarios tienen los derechos para usar, cambiar y redistribuir los datos. También permite buscar en textos completos y dispone de una API para todas las funciones de búsqueda. Visualización: Es una herramienta que ofrece una variedad de formas para ver y procesar los datos, tales como gráficos, vistas de tabla, datos de mapeos, datos de imagen. El sistema operativo utiliza una extensión CKAN DataStore, planteándose como una base de datos que permite almacenar y proveer, además, una API que facilita la integración con otras áreas de la plataforma. Metadatos: Proporciona una amplia variedad de metadatos para los conjuntos de datos, múltiples formatos si se proporcionan e incluye además una API. Interfaz muy personalizables y amigables (OKF. Open Knowledge Foundation, 2020).
--	--

3.5.4. INVENIO

	INVENIO https://invenio-software.org/
Origen	2016-2019: Se desarrollan 3 productos: InvenioRDM, InvenioILS e InvenioFramework. En 2020 se lanzará InvenioRDM, una aplicación desarrollada para repositorios. CERN (Organización Europea para la Investigación Nuclear) y colaboradores desarrollan estos productos con la filosofía de Ciencia Abierta.

Objetivo	-InvenioRDM tiene un rol relevante en el ecosistema de investigación actual para difundir y preservar, permitir la reproducibilidad y potenciar la reutilización.
Experiencias	Depósito General de Documentos (Barcelona), EDUDOC (Suiza), Zenodo (datos y biblioteca), Datos Abiertos del CERN (datos), HEPdata (datos), B2SHARE (datos), IDEP, CDSVideos (CERN).
Características	Proyecto colaborativo que ofrece tres productos a la comunidad: InvenioRDM: plataforma de gestión de documentos-repositorio; InvenioIIS: sistema integrado de biblioteca e Invenio Framework: biblioteca de códigos para construir sistemas de información a gran escala como InvenioRDM e InvenioIIS. Nos centraremos en las características de InvenioRDM. Es un software gratuito de código abierto, al igual que una aplicación para repositorios y se diferencia de Zenodo, el cual se define como un servicio. UX interfaz: Declaran la mejor interfaz para usuarios finales, curadores, administradores de sistema y desarrolladores. Integración institucional: Para esto, admite su solución de inicio de sesión único (SSO), sus sistemas de almacenamiento, sus nubes públicas / privadas, sus modelos de permisos, así como sus campos personalizados. Comunidades de usuarios a través de perfiles: Los incluye como preconfigurados para repositorios institucionales, sistemas de gestión de datos de investigación (RDM), así como repositorios específicos de dominio para ciencias biomédicas y de salud. Escalable manejar de 1 a 100 millones de registros, 1 byte o varios petabytes. NGR (Repositorio de Próxima Generación): Implementa esta visión para así integrarse con otras infraestructuras de Ciencia Abierta como ORCID, DataCite y OpenAIRE. Seguridad y preservación a largo plazo. Arquitectura web moderna y estándares que facilitan la implementación, el mantenimiento y el uso. InvenioRDM se está desarrollando con una amplia gama de características para optimizar las buenas prácticas de datos y aumentar el valor durante todo el ciclo de vida de la investigación. Es nativo de JSON y proporciona API RESTful listas para usar, que permitirán crear aplicaciones encima del repositorio. Elasticsearch: Es un motor de búsqueda que soporta todo tipo de consultas, desde texto a geoespacial. Dispone de dos bases de datos relacionales PostgreSQL y MySQL y fue desarrollado utilizando Python 3. (INVENIO, © 2016-2019).

3.5.5. FEDORA

	FEDORA (Flexible Extensible Digital Object Repository Architecture) https://duraspace.org/fedora/
Origen	1998: Se desarrolla por un Grupo de Investigación de la Biblioteca Digital, en el Departamento de Informática de la Universidad de Cornell.⁶⁹ 2020: Actualmente, se encuentra liderado por el Grupo de Liderazgo Fedora y bajo la dirección de DuraSpace, una organización sin fines de lucro.
Objetivo	-Convertir al Repositorio en un componente de FEDORA, permitiéndole proporcionar servicios para depositar, almacenar, replicar y acceder a DigitalObjects como entidades genéricas, identificables por sus nombres únicos (URN).
Experiencias	Repositorio de datos de DUKE Research, CLARIN-D de Stuttgart, FORDS DARIS (Suiza), CIESIN (Estados Unidos), ARCHE (Austria), 4TU.ResearchData (Países Bajos), SEDAC (Estados Unidos).
Características	Fedora es la plataforma de repositorio flexible, modular y de código abierto⁷⁰ con soporte nativo de datos vinculados para la gestión y difusión de contenido digital. Recomendable para bibliotecas y archivos digitales, tanto para acceso como para preservación. Fedora es un componente clave de marcos de aplicaciones más grandes al estilo de Samvera⁷¹ e Islandora⁷². API RESTful⁷³, protocolos OAI-PMH⁷⁴ y protocolo SWORD (interoperabilidad): Pueden almacenar, preservar y proporcionar acceso a cualquier tipo de archivo. Soporte de datos vinculados nativos, porque es un servidor del Linked Data Platform⁷⁵. Perfiles seguros en base a roles, XACML (eXtensible Access Control Markup Language) o control de acceso web.

⁶⁹ Flexible Extensible Digital Object Repository Architecture (FEDORA). Es una plataforma innovadora, gratuita y de código abierto para hardware, nubes y contenedores, que permite a los desarrolladores de software y miembros de la comunidad construir soluciones personalizadas para sus usuarios. Sitio web:

<http://web.archive.org/web/20000819130403/www.cs.cornell.edu/cdlrg/FEDORA.html>

⁷⁰ Licencia. Sitio web: <https://duraspace.org/fedora/resources/fedora-open-source-software-license/>

⁷¹ Samvera es un marco de repositorio de código abierto (comunidad internacional). Sitio web: <https://samvera.org/>

⁷² Islandora es un marco de repositorio de código abierto (Canadá). Sitio web: <https://islandora.ca/>

⁷³ Una API RESTful (también conocida como servicio web RESTful o RESTful API) se basa en la transferencia de estado de representación (REST), un estilo arquitectónico y un enfoque de las comunicaciones que a menudo se utilizan en el desarrollo de servicios web. Sitio web: <https://searchapparchitecture.techtarget.com/definition/RESTful-API>

⁷⁴ OAI-PMH es un protocolo de Iniciativa de Archivos Abiertos para la recolección de metadatos.

⁷⁵ LDP define un conjunto de reglas para operaciones HTTP en recursos web, algunas basadas en el Marco de Descripción de Recursos (Resource Description Framework, RDF) para proporcionar una arquitectura destinada a leer y a escribir datos vinculados en la web. Sitio web: <https://www.w3.org/TR/ldp/>

	Escalable: Puede manejar millones de archivos y registros de metadatos. Preservación, para lo cual proporciona la comprobación de la fijación, un seguimiento de auditoría, el control de versiones y la importación/exportación. Flujos de trabajo. Esquema de metadatos Dublin Core y METS. Uso de identificadores persistentes como DOI y URIs. Sistema operativo UNIX, Mac y Windows. (Fedora, 2020).
--	---

3.5.6. DSPACE CRIS

	DSPACE CRIS (Current Research Information Systems) https://wiki.lyrasis.org/display/DSPACECRIS/DSpace-CRIS+Home
Origen	4Science-DURASPCE-DSpace.
Objetivo	-Permitir la ingestión, almacenamiento, visualización y gestión de datos y metadatos para todas las entidades de investigación que se consideren.
Experiencias	Durban University of Technology (Sudáfrica), Proyecto Nacional PeruCRIS (Perú, proyecto en curso), Universidad de Winnipeg (Canadá), Administración Nacional de Laboratorios e Instituto de Salud “Carlos G Malbrán”. Sistema de Gestión del Conocimiento (Argentina), CEDES (Argentina) y la Pontificia Universidad Católica Argentina, Universidad de Nueva Inglaterra (Australia), Universidad de Hong Kong, Universidad de Singapur, euroCRIS (Países Bajos), Universidad Tecnológica de Chipre.
Características	Extensión gratuita de código abierto, porque amplía la funcionalidad del software DSpace y expande su modelo de datos mientras permanece alineado con su base de código. Estándares y prácticas internacionales, tales como: ORCID API v.2 (cumplimiento completo que incluye extracción/inserción de información para perfiles, publicaciones, proyectos). Se está lanzando el cumplimiento de la API ORCID v3. Señalización y sincronización de recursos (que implementan los comportamientos recomendados de COAR NGR⁷⁶). Pautas de OpenAIRE para gestores de literatura, para archivos de datos, para CRIS Managers (basado en CERIF⁷⁷ nov. 2019).

⁷⁶ Next Generation Repositories (NGR) son repositorios de Próxima Generación.

⁷⁷ Common European Research Information Format (CERIF). Se trata de un estándar que la UE recomienda a sus estados miembros para registrar información sobre la actividad de investigación, que fue desarrollado euroCRIS, que actualmente lo mantiene también. Sitio web: <https://www.dcc.ac.uk/>

	Planes de AA completo e inmediato, acorde con los principios del Plan S que emana de la Coalición S⁷⁸. Principios FAIR. Modelo de datos flexible permite recopilar y administrar datos e información de investigación típicos de un sistema CRIS (potencia la interoperabilidad), para definir entidades y atributos con sus enlaces recíprocos. Si solo desea mejorar la gestión de los autores, proporcionar variantes de nombre e identificaciones como ORCID, explotar el variado ecosistema de identificadores persistentes, vincular investigadores con proyectos, premios, etc. Evaluación, informes y análisis, proporciona gráficos de red de colaboración, bibliometría y estadísticas agregadas (por investigador, por departamento) con informes gráficos, CV y bibliografías, integración con ORCID API v.3, citas y enlaces directos a Google Scholar y Almetrics. Identificadores persistentes múltiples DOI, Handle, URN, así como ISBN, ISSN, URL, otros. Integraciones con complementos: Integración CKAN - Gestión de datos de investigación (Donohue, s.f.)
--	--

3.6. Principios FAIR

En el siguiente apartado, me referiré a los llamados Principios FAIR, a través de cuyo nombre se alude a las características que deben cumplir los datos de investigación, entendiendo por tales no sólo los datos duros y numéricos, sino también fotografías, infografías, gráficos y audios; es decir, toda información en que aquélla se sustenta. Estos principios rectores provienen del inglés y son: Findable o Encontrable, Accessible o Accesible, Interoperable o Interoperable y Reusable o Reutilizable, constituyéndose así en el sustrato del AA y de la CA.

⁷⁸ La Coalición S está conformada en la actualidad por 24 integrantes, entre los que se cuentan la Organización Mundial de la Salud, la Comisión Europea, el Consejo Europeo de Investigación, la United Kingdom Research and Innovation (UKRI), la Fundación Bill y Melinda Gates, el Programa para la Investigación y Entrenamiento en Enfermedades Tropicales y el Programa de de las Naciones Unidas para el Desarrollo (UNDP), entre otras. Una de sus iniciativas es el Plan S, que promueve el Acceso Abierto de todas las publicaciones científicas, lo que ha generado debate especialmente con las editoriales y revistas especializadas. Sitio web:

[https://revistas.javeriana.edu.co/files-articulos/UPSY/19%20\(2020\)/64762919001/](https://revistas.javeriana.edu.co/files-articulos/UPSY/19%20(2020)/64762919001/)

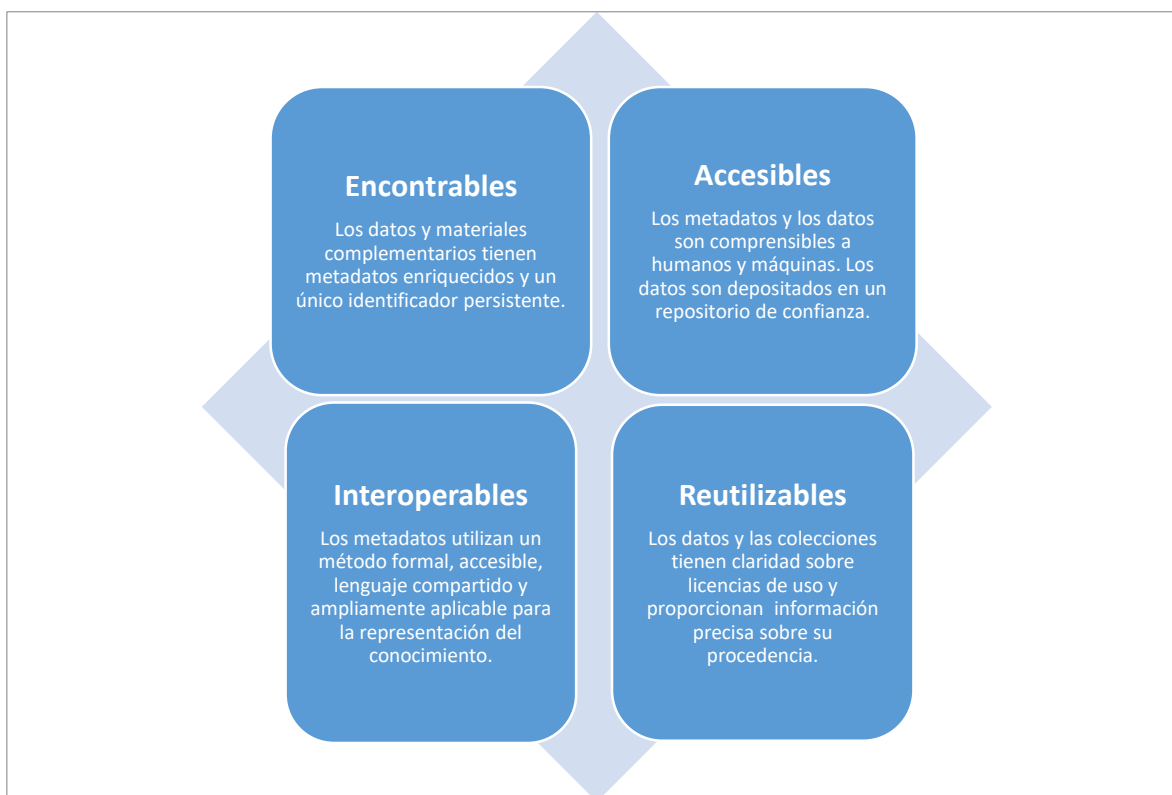


Figura 3-10: Principios FAIR

Adaptado de ¿Qué son los Principios FAIR?, obtenido de:

<https://libereurope.eu/wp-content/uploads/2017/12/LIBER-FAIR-Data.pdf>

3.6.1. ¿Por qué son importantes los Principios FAIR?

Los Principios FAIR son importantes porque corresponden a iniciativas para encontrar, acceder, compartir y reutilizar los resultados de la investigación, obtenidos de organizaciones o agencias de financiación, las cuales respaldan fuertemente a la CA. Los Principios FAIR permiten fortalecer el progreso de los descubrimientos. Para que esto ocurra, se requiere del desarrollo e implementación de infraestructuras y servicios orientados a estándares donde prevalezcan las buenas prácticas. En dicho contexto, los Principios FAIR fortalecen esta aspiración como se ha mencionado.

Para abordar este tema, me basaré en el artículo Los principios rectores FAIR para la gestión y administración de datos científicos (Wilkinson, 2016).

El artículo de Wilkinson (con una creciente métrica de menciones, accesos y citas) enfatiza que los Principios FAIR ponen “un énfasis específico en mejorar la capacidad de las máquinas para encontrar y utilizar automáticamente los datos, además de apoyar su reutilización por parte de las personas”. Más adelante, describe los cuatro principios fundamentales: capacidad de búsqueda, accesibilidad, interoperabilidad y reutilización, con el propósito de “guiar a los productores y editores de datos mientras navegan por estos obstáculos, lo que ayuda a maximizar el valor agregado obtenido por la

publicación académica digital formal contemporánea”, considerando además el anhelo por parte de los autores en cuanto a que estos cuatro principios se extienda, además, a los algoritmos, herramientas y flujos de trabajo que originaron esos datos (Wilkinson, 2016).

3.6.2. ¿Qué inspira el surgimiento de estos Principios?

En el ciclo de vida de los datos de investigación, los repositorios tienen un rol y algunos de éstos han sido implementados con robustas tecnologías y buenas prácticas, permitiendo que sus datos sean un aporte frente a una necesidad. Sin embargo, el ecosistema de datos en expansión es poco integrado y variado, lo que repercute desfavorablemente en el descubrimiento y en la reutilización de dichos datos no sólo por parte de los humanos, sino también de las máquinas.

Una importante iniciativa se dio en 2014, en la localidad de Leiden (Países Bajos), con la realización del taller “Jointly Designing a Data Fairport”, actividad que contó con una alta asistencia y que concluyó con el interés por mejorar el descubrimiento de datos y su reutilización, a través de la definición y el apoyo generalizado de un conjunto mínimo de principios y prácticas orientadoras. Todas aquellas iniciativas fueron acordadas por la comunidad, permitiéndoseles a las partes la facultad de descubrir, acceder, integrar y reutilizar fácilmente los contenidos, incluidas las citaciones.

De este encuentro surgió un borrador con un conjunto de principios, el cual -gracias al posterior trabajo mancomunado de un grupo de profesionales en pro de su perfeccionamiento- dio origen a los Principios que hoy están cada vez más presentes y deben ser considerados al momento de implementar un repositorio.

El artículo releva la importancia de las máquinas en contextos de investigación donde proliferan datos. Si bien una máquina no tiene la capacidad para discernir la obtención y análisis de datos, supera la capacidad del ser humano en cuanto a no operar al alcance, la escala y la velocidad que requieren la actual producción exponencial de los datos. Según los autores, “es por esta razón que los seres humanos dependen cada vez más de agentes computacionales para realizar tareas de descubrimiento e integración en su nombre. Esto requiere que las máquinas sean capaces de actuar de manera autónoma y apropiada cuando se enfrenten a la amplia gama de tipos, formatos y mecanismos/protocolos de acceso que se encontrarán durante su exploración autoguiada del ecosistema de datos global. También requiere que las máquinas mantengan un registro exquisito de procedencia, de modo que los datos que están recopilando puedan citarse de manera precisa y adecuada” (Wilkinson, 2016).

Este texto alude al concepto “accionable por máquina” refiriéndose a un continuo de eventuales estados, donde un objeto digital entrega información al momento de ejecutarse un descubrimiento. Frente a dicho escenario, una máquina debería:

- Identificar el tipo de objeto (con respecto tanto a la estructura como a la intención).
- Determinar si es útil dentro del contexto de la tarea actual del agente al interrogar metadatos y/o elementos de datos.
- Determinar si es utilizable, con respecto a la licencia, consentimiento u otras restricciones de accesibilidad o uso.
- Tomar las medidas apropiadas, de la misma manera que lo haría un humano (Wilkinson, 2016).

Al referirse a que sea accionable por máquina, los autores expresan su aplicación en dos situaciones que pueden ser abordadas por máquinas y que son, además, propiciadas por la aplicación de los Principios FAIR:

1° Cuando se refiere a los metadatos contextuales que rodean un objeto digital (¿qué es?).

2° Cuando se refiere al contenido del objeto digital en sí mismo (¿cómo lo proceso o cómo lo integro?).

El artículo incorpora una versión mejorada tomada del borrador, la que se comparte a continuación en una versión adaptada en la Tabla 1.

Tabla 3-1: Principios rectores FAIR

Para ser Localizable:	F1 (meta)datos⁷⁹ se les asigna un identificador globalmente único y persistente F2 los datos se describen con metadatos enriquecidos (definidos por R1 a continuación) F3 los metadatos incluyen clara y explícitamente el identificador de los datos que describe F4 (meta) los datos se registran o indexan en un recurso de búsqueda
Para ser Accesible:	A1 (meta) los datos se pueden recuperar por su identificador utilizando un protocolo de comunicaciones estandarizado A1.1 el protocolo es abierto, gratuito y universalmente se puede implementar A1.2 el protocolo permite un procedimiento de autenticación y autorización, cuando sea necesario A2 los metadatos son accesibles, incluso cuando los datos ya no están disponibles

⁷⁹ Se utiliza cuando el Principio debe aplicarse tanto a los metadatos como a los datos.

Para ser Interoperable:	I1 (meta) los datos usan un lenguaje formal, accesible, compartido y ampliamente aplicable para la representación del conocimiento I2 (meta) datos usan vocabularios que siguen los principios FAIR I3 (meta) los datos incluyen referencias calificadas a otros (meta) datos
Para ser Reutilizable:	R1 describir meta(datos) detalladamente con una pluralidad de atributos precisos y relevantes R1.1 (meta) los datos se liberan con una licencia de uso de datos clara y accesible R1.2 (meta) los datos están asociados con una procedencia detallada R1.3 (meta) los datos cumplen con los estándares de la comunidad relevantes para el dominio

Según los autores, los Principios FAIR se encuentran vinculados, aunque son separables e independientes y siempre se debería considerar definir mínimamente cada uno de ellos. También se pueden lograr en cualquier combinación entre ellos y se pueden iniciar de modo incremental.

En el artículo también se aborda la diferencia entre datos y metadatos para referirse a situaciones especiales, como son -dentro de éstas- los datos sensibles. La publicación de metadatos enriquecidos para facilitar el descubrimiento, incluidas reglas claras con respecto al proceso de acceso a los datos, proporciona un alto grado de FAIRness, es decir, no deberían ser describibles, descubribles y reutilizables, obviando los principios.

Además, es importante considerar que los autores sostienen que los Principios FAIR preceden a la implementación. Tampoco sugieren una tecnología estándar o solución de implementación particular. Agregan que estos principios no son en sí mismos una especificación o estándar, aclarando que su propósito es actuar como una guía para editores y administradores de datos que requieren que sus flujos de investigación sean Encontrados, Accesibles, Interoperables y Reutilizables (Wilkinson, 2016).



Figura 3-11: Sitio web de los Principios FAIR

Fuente: <https://www.go-fair.org/fair-principles/>

Ejemplos para cada uno de los principios:

- Localizable: Se puede lograr con DOI un identificador persistente, como consecuencia esto permite cita.
- Accesible: Es importante la asignación de metadatos.
- Interoperable: Junto con los metadatos, se recomienda el uso de las API's estándares, ontologías y vocabularios controlados.
- Reutilizable. Se recomienda la iniciativa Data Documentation Initiative (DDI)⁸⁰, así como declarar las licencias y términos empleados.

3.7. Certificaciones

¿Cómo lograr que un repositorio sea confiable, en el sentido de que los investigadores que depositan ahí sus datos tengan la tranquilidad de que serán correctamente identificados y preservados?

⁸⁰ Data Documentation Initiative (DDI) es un estándar gratuito internacional para documentar y administrar diferentes etapas en el ciclo de vida de los datos de investigación, como la conceptualización, la recopilación, el procesamiento, la distribución y el descubrimiento, permitiendo describir los datos producidos por encuestas y por otros métodos de observación en las ciencias sociales, conductuales, económicas y de salud. Sitio web: <https://ddialliance.org/>

Esta interrogante, al igual que otras en este mismo sentido, rondaron el inicio de la colaboración entre Reserach Library Group (RLG) y Online Computer Library Center (OCLC), en 2000, para establecer en conjunto atributos respecto de las características y de las responsabilidades para hacer de un repositorio un recurso confiable.

Por otra parte, el Center for Reserach Library (CRL)⁸¹ empezó en 2006 a certificar repositorios de interés para su comunidad. Bajo la modalidad Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRACT) certificó, entre 2010 y 2018, seis repositorios, cuales son: Canadiana.org, Chronopolis, CLOCKSS, Hathitrust, Portico y Scholars Portal (CRL. Center for Research Libraries Global Resources Network, s.f).

En tanto, las métricas de preservación que CRL aplica permiten mantener los recursos digitales en el transcurso del tiempo, lo cual -además de brindar tranquilidad a los investigadores- posibilita el emprendimiento de nuevas investigaciones o estudios longitudinales.

Existen varias herramientas de evaluación para emplear métricas y comprobar repositorios confiables, de las cuales CRL aplica preferentemente:

- **TRACT** Auditoría y Certificación de Repositorios Confiables: Criterios y Lista de Verificación. TRACT es una lista de verificación de archivo del repositorio de confianza. Esta herramienta contiene métricas que permiten evaluar un repositorio en las áreas de administración y política, manejo de objetos y tecnología, equivalentes a “Estructura organizacional”, “Gestión de objetos digitales” y “Tecnologías, infraestructura tecnológica y seguridad”.
- **ISO 16363: 2012**⁸². Se le conoce también como Lista de Verificación del Repositorio Digital de Confianza (TDR)⁸³. Corresponde a una revisión de la lista de verificación de TRACT.

Adicionalmente, CRL aplica otros elementos para la evaluación:

- **Diez Principios:** estos se acordaron en colaboración con tres instituciones conocedoras de temas de preservación digital: The Digital Curation Center (DCC), en Reino Unido; DigitalPreservationEurope, NESTOR, en Alemania; y el CRL, en Estados Unidos. En conjunto, se acordaron diez principios básicos para la auditoría y certificación de repositorios (CRL. Center for Research Libraries. Global Resources Network, s.f).

Asimismo, el CRL proporciona otras herramientas de evaluación para la certificación de repositorios, de las cuales se destacan:

⁸¹ Center for Reserach Library (CRL) se fundó en 1949, en Estados Unidos, definiéndose como un consorcio internacional de bibliotecas universitarias e independientes de investigación, destinado a apoyar la investigación original y la enseñanza inspirada en las humanidades, las ciencias y las ciencias sociales, preservando y poniendo a disposición de los académicos una gran cantidad de materiales de fuentes primarias raras y poco comunes de todas las regiones del mundo. Sitio web: <https://www.crl.edu/about>

⁸² ISO1663:2012. Sitio web: <https://www.iso.org/standard/56510.html>

⁸³ Trustworthy Digital Repositories (TDR). Sitio web: <https://public.ccsds.org/pubs/652x0m1.pdf>

- **CoreTrustSeal**⁸⁴. Esta certificación se concibe como el primer paso en un marco global para la certificación de un repositorio, incluyendo la certificación de nivel extendido (Nestor-Seal DIN 31644) y la certificación de nivel formal (ISO 16363). Los revisores (voluntarios calificados) realizan las auditorías, mientras que un Comité Asesor asiste a la Junta de CoreTrustSeal. Por este trabajo, CoreTrustSeal cobra una tarifa modesta para cubrir los costos administrativos. Esta certificación reemplaza a otras certificaciones y, actualmente, goza de mucha popularidad.
- **Data Asset Framework (DAF)**⁸⁵. Creada en el Reino Unido en colaboración con la Universidad de Glasgow, DCC y JISC, también es utilizada como una métrica para auditoría de evaluación de repositorios.
- **Digital Asset Assessment Tool (DAAT)**⁸⁶. Fue parte del JISC (Joint Information Systems Committee) y, en la actualidad, es un proyecto ya finalizado que incluyó una evaluación de herramientas de prueba de formato archivo.
- **Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)**⁸⁷. Entrega una metodología para la autoevaluación, alentando a las organizaciones a establecer una autoconciencia integral de sus objetivos, actividades y activos antes de identificar, evaluar y gestionar los riesgos implícitos dentro de su organización.

Como contraparte a Trusted Repositories Audit & Certification (TRAC), se dispone de la European Framework for Audit and Certification of Digital Repositories⁸⁸, la cual propone tres niveles para la certificación: Core, Extended y Formal, también conocidos como Bronce, Plata y Oro. Las principales áreas para evaluación son las mismas en estos tres niveles y se orientan a la organización, al manejo de entidades intelectuales y a las representaciones e infraestructura y seguridad. Un alcance a cada una de estas certificaciones corresponde a:

- Certificación Básica (CORE): Se otorga a los repositorios que obtienen la certificación DSA⁸⁹.
- Certificación Extendida (EXTENDED): Se otorga a los repositorios de certificación básica que, además, realizan una auto auditoría estructurada, revisada externamente y disponible de forma pública, basada en la norma ISO 16363⁹⁰ o en la norma DIN 31644.

⁸⁴ CoreTrustSeal. Sitio web: <https://www.coretrustseal.org/about/>

⁸⁵ Data Asset Framework (DAF). Sitio web: <https://www.data-audit.eu/>

⁸⁶ Digital Asset Assessment (DAAT). Sitio web: <https://www.webarchive.org.uk/wayback/archive/20140614062318/http://www.jisc.ac.uk/whatwedo/programmes/preservation/assetmanagement/daat.aspx>

⁸⁷ Digital Repository Audit Method Based on Risk Assessment (DRAMBORA). Sitio web: <http://www.repositoryaudit.eu/>

⁸⁸ European Framework for Audit and Certification of Digital Repositories (Marco Europeo de Auditoría y Certificación de Repositorios Digitales, en español). Sitio web: <http://www.trusteddigitalrepository.eu/Welcome.html>

⁸⁹ Digital Signature Algorithm (DSA o Algoritmo de Firma Digital, en español). Es un estándar del Gobierno Federal de los Estados Unidos de América o FIPS para firmas digitales.

⁹⁰ ISO 16363: 2017. Sitio web: <https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma?c=N0058850>

- Certificación Formal (FORMAL): Se otorga a los repositorios que, aparte de la certificación básica, obtienen una auditoría externa completa y una certificación basada en la norma ISO 16363 o en la norma DIN 31644⁹¹ equivalente.

La tabla 2 resume los tres niveles y las diferencias entre sus certificaciones o alcances:

Tabla 3-2: Resumen de los tres niveles y la diferencia entre sus certificaciones

■ Security

The differences reside in how the audit is performed and the number of factors evaluated in that audit.

The following table summarizes the three levels and how their certifications differ:

LEVEL	CORE	EXTENDED	FORMAL
Organization(s)	WDS: ICSU World Data System DSA: Data Seal of Approval	DIN  : German Institute for Standardization	ISO  : International Organization for Standardization
No. of Requirements	16	34	100+
Standards	Mandatory Requirements 	DIN 31664 	ISO 14721 (OAIS)  ISO 16363  ISO 16919  ISO 17021 
Audit Process	Self-assessment + independent peer review (2)	Self-assessment + independent peer review (2)	ISO certified audit with accredited auditors
Certification Cost	Free	€500	\$10,000
Designation	World Data System logos or Data Seal of Approval	nestor Seal for Trustworthy Digital Archives	TBD
Certification lifespan	3 years	Indefinite	3 years
No. of Certified Repositories	130+ (WDS  DSA )	2 	Coming Soon

Ultimately, the choice of certification depends on how much a repository is willing to invest in its perceived

Fuente: https://datascience.nih.gov/trusted_digital_repository

Se puede apreciar que la elección de la certificación considera varios aspectos, pasando desde los presupuestarios hasta la aplicación de buenas prácticas adoptadas al momento de la implementación y gestión. En esta tabla, el autor (Lin D. , s.f.) proporciona un aproximado respecto de los costos de certificación, incluyendo los costos de las certificaciones en sí mismas, así como los costos del tiempo dedicado a esta labor.

Por otra parte, este artículo también menciona que la vida útil de una certificación es variable. En cuanto al nivel extendido de certificación (Extended), tiene una validez indefinida, pero debe actualizarse para mantener su relevancia. Mientras, los niveles Core y Formal son válidos sólo por tres años, respectivamente. También se debe considerar que los cambios tecnológicos, del mismo modo que las migraciones y las nuevas necesidades de los usuarios o de los establecimientos de políticas, pueden gatillar nuevas necesidades de recertificación (Lin D. , s.f.).

En relación a la preservación de datos, uno de los estándares frecuentemente utilizados para preservación digital se encuentra en el modelo de referencia Open Archival Information System

⁹¹ Digital Preservation Coalition (DPC). Sitio web: <https://www.dpconline.org/handbook/institutional-strategies/audit-and-certification>

(OAIS), el cual fue publicado como una norma internacional, denominada en su versión actual como ISO 14721:2012⁹².

Este estándar se puede definir como una referencia, la cual describe generalidades respecto de funciones, responsabilidades y la organización de un repositorio o sistema destinado a preservar datos digitales en el largo plazo.

La certificación es un componente que debe considerarse al momento de implementar o gestionar un repositorio de datos de investigación, lo cual proporcionará confiabilidad y acceso sostenible a los datos de investigación.

3.8. Repositorios de Próxima Generación

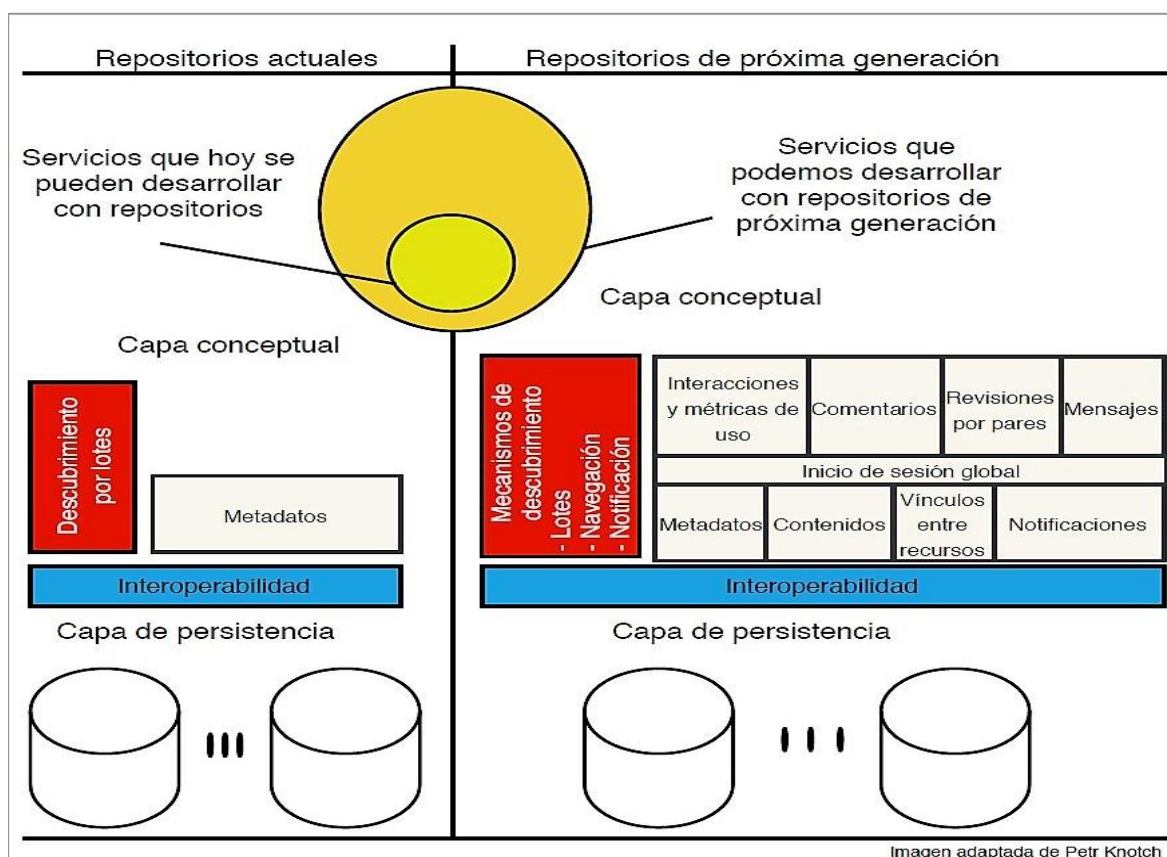


Figura 3-12: Repositorios actuales vs Repositorios de Próxima Generación

Tomado de COAR: Next Generation Repositories 2017 p. 13: <https://www.coar-repositories.org/files/NGR-Final-Formatted-Report-cc.pdf>

⁹² Sitio web: <http://www.oais.info/>

En noviembre de 2017, COAR⁹³ publicó el primer informe de Repositorios de Próxima Generación (NGR), que contiene una lista de diecinueve tecnologías y protocolos para sistemas de repositorio. Las recomendaciones se basaron en una amplia gama de testimonios de usuarios y comportamientos que fueron examinados y priorizados por la comunidad del repositorio (COAR. Building a Global Knowledge Commons, 2017).

Este documento se sustenta en el interés de COAR para trabajar colaborativamente con sus afiliados, con el propósito de que se puedan adoptar recomendaciones en los principales softwares y plataformas de código abierto para generar perfiles y servicios con valor agregado.

Se pretende optimizar el valor del ecosistema de repositorios para dotarlos de mejoras tecnológicas, las cuales permitan ampliar sus roles y la interoperabilidad brindada en la web. En dicho sentido, la visión y los aspectos técnicos que rodean los NGR se enfocan tanto en los servicios como en la infraestructura, propiciando fortalecer una arquitectura web donde se puedan identificar, evitando lo que ocurre actualmente, que muchos metadatos descriptivos se pierden o no se establecen las relaciones esperadas entre ellos.

Para subsanar lo anterior, en su necesidad de fortalecer las infraestructuras de los repositorios, COAR propicia los denominados Repositorios de Próxima Generación (NGR, en inglés)⁹⁴ para lo cual se ha trazado los siguientes objetivos:

- “Lograr un nivel de interoperabilidad entre repositorios exponiendo comportamientos uniformes en todos los repositorios que aprovechan las tecnologías y arquitecturas compatibles con la web, y mediante la integración con las infraestructuras académicas globales existentes, específicamente aquellas destinadas a la identificación de, por ejemplo, contribuciones, datos de investigación, contribuyentes, instituciones, financiadores, proyectos”.
- “Fomentar la aparición de servicios de valor agregado que utilizan estos comportamientos uniformes para respaldar el descubrimiento, el acceso, la anotación, la curación en tiempo real, el intercambio, la evaluación de la calidad, la transferencia de contenido, el análisis, el rastreo de procedencia, etc.”
- “Ayudar a transformar el sistema de comunicación académico al enfatizar los beneficios de la gestión colectiva, abierta y distribuida, el contenido abierto, los comportamientos uniformes, la difusión en tiempo real y la innovación colectiva” (COAR. Confederation Open Access Repositories, s.f).

⁹³ Confederation of Open Access Repositories (COAR). Se trata de una asociación internacional que reúne repositorios individuales y redes de repositorios para crear capacidad, alinear políticas y prácticas y actuar como una voz global para la comunidad de repositorios. Actualmente, cuenta con 157 integrantes y socios provenientes de 53 países, que representan bibliotecas, universidades, instituciones de investigación, financiadores gubernamentales y otros. En Chile, ANID y LA Referencia.

⁹⁴ Next Generation Repositories (NGR). Sitio web: <https://universoabierto.org/2017/11/30/la-proxima-generacion-de-repositorios-de-acceso-abierto-recomendaciones-tecnicas-de-coar/>

Para alcanzar estos objetivos, se propone una serie de herramientas tecnológicas que incluyen estándares y protocolos que deberían ser adoptados.

El Grupo de Trabajo de NGR se ha enfocado, explícitamente, en tecnologías genéricas requeridas por todos los repositorios con el propósito de apoyar la adopción de comportamientos comunes. Sin embargo, también reconocemos que existen otras tecnologías y estándares que pueden ser útiles para ciertos tipos de contenido o comunidades disciplinarias.

En algunos casos, las tecnologías requeridas para soportar un comportamiento específico aún no están lo suficientemente maduras, o aún no está claro qué tecnología prevalecerá. En otros casos, incluso, no hay tecnologías apropiadas en la actualidad para soportar el comportamiento específico. Casos en los cuales este Grupo de Trabajo no puede recomendar tecnologías específicas, aunque declara que continúan monitoreando los desarrollos y haciendo recomendaciones en la medida que surjan nuevas tecnologías o mejoren las disponibles (COAR. Confederation Open Access Repositories, s.f).

3.9. Tecnologías⁹⁵

- ✓ Activity Streams 2.0
- ✓ COUNTER
- ✓ Creative Commons Licenses
- ✓ ETag
- ✓ HTTP Signatures
- ✓ International Image Interoperability Framework
- ✓ IPFS
- ✓ Linked Data Notifications
- ✓ OpenID Connect
- ✓ ORCID
- ✓ ResourceSync
- ✓ Signposting
- ✓ Sitemaps
- ✓ Social Network Identities
- ✓ SUSHI
- ✓ SWORD
- ✓ Web Annotation Model and Protocol
- ✓ WebID
- ✓ WebID/TLS
- ✓ Webmention
- ✓ WebSub

⁹⁵ Detalle de las tecnologías para NGR en: <https://ngr.coar-repositories.org/technology/>

Actualmente, los softwares INVENIO y DSpace CRIS son considerados por sus infraestructuras, mientras que los NGR son considerados por sus servicios.

3.10. Repositorios de datos de investigación en Europa

El estado del arte de los repositorios de datos abiertos en Europa tiene su referente a través del Portal Europeo de Datos, cuya decisión de creación en la Unión Europea (UE) fue establecida en 2012 por medio de un documento oficial relativo a la reutilización de los documentos de la Comisión, en el cual quedó considerado en particular su artículo 10⁹⁶ referente a la creación de un portal de datos (EU. European Union, 2020). En tanto, la versión beta se lanzó en el segundo semestre de 2015, denominándose Portal Europeo de Datos, para lo cual se invitó a los 26 países integrantes y a las cinco instituciones internacionales constitutivas de la UE a colocar sus datos a disposición del público, siempre que fuese posible. Esto significa que los datos pueden reutilizarse de forma gratuita y sin restricciones de derechos de autor (EU. European Union, 2020).

Esta resolución permitiría a la población de las instituciones participantes tener en perspectiva los potenciales beneficios que dicha decisión pudiese aportar para el desarrollo económico y para la transparencia de los estados e instituciones de la UE.

Si bien el portal consideró datos abiertos tales como datos geográficos, geopolíticos, financieros, estadísticas, resultados electorales y actos jurídicos, también fueron incluidos datos sobre criminalidad, salud, medio ambiente, transporte e investigación científica.

Según consta, estos datos son de libre acceso y pueden reutilizarse en bases de datos, informes o proyectos. Las instituciones y organismos de la UE proporcionan estos datos en distintos formatos.

Por repositorios de datos de investigación entenderemos, entonces, “un sistema de información técnica y organizativa que ayuda a los investigadores a administrar, almacenar y proporcionar sus propios conjuntos de datos, y a encontrar y acceder fácilmente a conjuntos de datos de otras fuentes” (Kindling, 2017). El acercamiento a su actual estado en Europa y a los NGR fue a través de varias vías, dentro de las cuales el directorio más representativo en esta temática es re3data.org, cuyo servicio tiene por objetivo “indexar y describir repositorios de datos de investigación para presentar información detallada sobre los servicios existentes” (Kindling, 2017).

⁹⁶ Debe crearse un portal de datos como punto de acceso único a los documentos disponibles para ser reutilizados. Además, entre estos documentos, conviene incluir la información sobre sus investigaciones elaborada por el Centro Común de Investigación. Se insta a adoptar una disposición que tenga en cuenta la tendencia hacia formatos legibles por máquina. Una importante mejora con respecto a la Decisión 2006/291/CE (Euratom) consiste en ofrecer -en general- los documentos de la Comisión para su reutilización a través de licencias abiertas de reutilización o sencillas cláusulas de exención de responsabilidad, sin necesidad de efectuar solicitudes individuales.

Desde 2017 a 2020 (dato tomado el 21 de junio de 2020) los repositorios de datos registrados en este directorio han experimentado un crecimiento total del orden de un 35%.

Registry of Research Data Repositories (más conocido como re3data.org, Registro de Repositorios de Datos de Investigación, en español) es un actual y cada vez más consolidado referente, el cual permite entregar un actualizado estado del arte de los repositorios de datos de investigación. Igualmente, está incursionando con importantes iniciativas de colaboración. Al respecto, cabe mencionar la Fundación Alemana de Investigación (DFG)⁹⁷, en funcionamiento desde enero de 2020, que comenzó a financiar el “Proyecto re3data - Referencia Abierta Dirigida por la Comunidad para Repositorios de Datos de Investigación (re3data - COREF)”⁹⁸, que tendrá un plazo de 36 meses para concretar un ambicioso proyecto conjunto de DataCite, el Instituto de Tecnología de Karlsruhe (KIT), la Escuela de Bibliotecas y Ciencias de la Información de Berlín en la Humboldt-Universität zu Berlin y la Oficina de Ciencias Abiertas de Helmholtz, en el Centro de Investigación de Geociencias de Alemania (GFZ)⁹⁹, en alemán).

El objetivo del proyecto es conectar re3data como referencia para repositorios de datos de investigación con otros servicios e infraestructuras. Para esto, re3data proporcionará descripciones de repositorio de núcleo personalizables y extensibles, que son identificables de manera persistente y pueden ser referidos y citados de manera apropiada. Esto incluirá una identificación única en la comunicación de (M2M) máquina a máquina. La plataforma técnica deberá ser avanzada para facilitar la reutilización de información confiable en repositorios de datos de investigación. Se incluirá, también, una API elaborada, así como herramientas y widgets fáciles de insertar.

Por otro lado, se ampliará el esquema de metadatos (en la actualidad, cuenta con 62) y mejorará vocabularios controlados. Todo esto para propiciar la calidad, la confianza y el cumplimiento de los Principios FAIR.¹⁰⁰

Para lograr estos ambiciosos objetivos y metas, el Consejo Editorial Internacional de re3data.org contribuirá al proyecto al igual que muchos socios externos. Mientras, con el propósito de integrar

⁹⁷ Deutsche Forschungsgemeinschaft (DFG o Sociedad Alemana de Investigación, en español). Sitio web: <https://www.dfg.de/>

⁹⁸ Detalles del Proyecto re3data - Referencia Abierta Dirigida por la Comunidad para Repositorios de Datos de Investigación (Community Driven Open Reference for Research Data Repositories, en inglés). Sitio web: <https://gepris.dfg.de/gepris/projekt/422587133?language=en> Consultado 210620.

⁹⁹ Sitio web: <https://www.gfz-potsdam.de/en/home/>

¹⁰⁰ Los principios FAIR fueron establecidos para la gestión y administración de datos, con la intención de proporcionar pautas para mejorar la capacidad de búsqueda, accesibilidad, interoperabilidad y reutilización de activos digitales. Sitio web: <https://www.go-fair.org/fair-principles/> Consultado 210620.

el servicio dentro del panorama de datos de investigación, re3data.org adoptará también las próximas recomendaciones de RDA¹⁰¹, EOSC¹⁰² y NFDI¹⁰³.

En consecuencia, “el proyecto conducirá a mejoras en la calidad general de los datos y proporcionará nuevos servicios adaptados precisamente a las necesidades de los usuarios y la comunidad”¹⁰⁴ (Cousijn H. y., 2020).

Todos los repositorios de datos de investigación registrados en este directorio (a la fecha, más de dos mil) proporcionan una descripción general del repositorio, incluyendo el nombre, la URL¹⁰⁵, el alcance de la temática y una descripción. Aquí se encuentran repositorios de datos de índole general o multidisciplinarios, al igual que especializados o temáticos.

Si una organización desea que su repositorio de datos sea indexado, un grupo editorial -al cual se sumó, además, DataCite¹⁰⁶- deberá previamente evaluarlo. En la sección Sugerencias de este mismo directorio, es posible encontrar el detalle de la política para registrar su repositorio de datos¹⁰⁷.

No se puede concebir la implementación de un repositorio de datos de investigación sin identificadores persistentes. Para el caso de re3data.org, cada registro es accesible y citable de manera persistente a través de un Identificador de Objeto Digital (DOI, en inglés)¹⁰⁸.

¹⁰¹ Research Data Alliance (RDA). Fue creada en 2013 por la Comisión Europea, la Fundación Nacional de Ciencias del Gobierno de los Estados Unidos, el Instituto Nacional de Estándares y Tecnología, y el Departamento de Innovación del Gobierno de Australia con el propósito de construir la infraestructura social y técnica para permitir el intercambio abierto y la reutilización de datos. Sitio web: <https://www.rd-alliance.org/about-rda> Consultado 210620.

¹⁰² Nube Europea de Ciencias Abiertas de la Comunidad Europea (European Open Science Cloud, EOSC). Es una de las 49 asociaciones europeas propuestas en Horizonte Europa, cuyo objetivo es acelerar el despliegue y la consolidación, en Europa, de un entorno abierto, confiable, virtual y federado con el propósito de almacenar, compartir y reutilizar datos de investigación. Expertos sostienen que será una “Red de datos FAIR y servicios relacionados para la ciencia”, mientras que OpenAIRE ha generado ya sus pautas para repositorios e interoperabilidad EOSC.

¹⁰³ Infraestructura Nacional de Datos de Investigación (Nationale Forschungsdateninfrastruktur). El Consejo Alemán recomendó, en abril de 2017, esta iniciativa porque el panorama de repositorios de datos alemán es complejo y altamente federado. En consecuencia, la NFDI deberá vincular, aumentar y complementar las infraestructuras ya existentes. Sitio web: <https://www.openaire.eu/data-repositories-germany> Consultado 210620.

¹⁰⁴ Consultado 210620.

¹⁰⁵ Uniform Resource Locator (URL o Localizador Uniforme de Recursos, en español). Es una secuencia de caracteres que sigue un estándar, permitiendo denominar recursos dentro del entorno de Internet para que puedan ser localizados.

¹⁰⁶ DataCite. Organización internacional sin fines de lucro que tiene como objetivo mejorar la cita de datos para establecer un acceso más fácil a los datos de investigación en Internet y aumentar la aceptación de los datos de investigación como contribuciones legítimas y citables al registro académico. Sitio web: <https://datacite.org/> Consultado 210620.

¹⁰⁷ re3data.org. Política de registro para repositorio. Sitio web: <https://www.re3data.org/suggest>

¹⁰⁸ Digital Object Identifier (DOI). Se trata de un Identificador de objeto digital de la IDF (International DOI Foundation), organización sin fines de lucro que es el organismo de Gobierno y gestión de la Federación de Agencias de Registro que proporciona servicios y registros para DOI, y es el registro modelo para el estándar ISO (ISO 26324), según el sistema DOI. El sistema DOI proporciona una infraestructura técnica y social para el registro y uso de identificadores interoperables persistentes, llamados DOI, para su uso en redes digitales.

Actualmente, destacados editores de revistas -entre los que se encuentran Copernicus Publications¹⁰⁹, PeerJ¹¹⁰, PLOS ONE¹¹¹ y Nature's Scientific Data¹¹²- recomiendan re3data.org en sus políticas editoriales, como una herramienta para que los autores depositen datos que respalden los resultados de investigación publicados en sus revistas (Kindling, 2017).

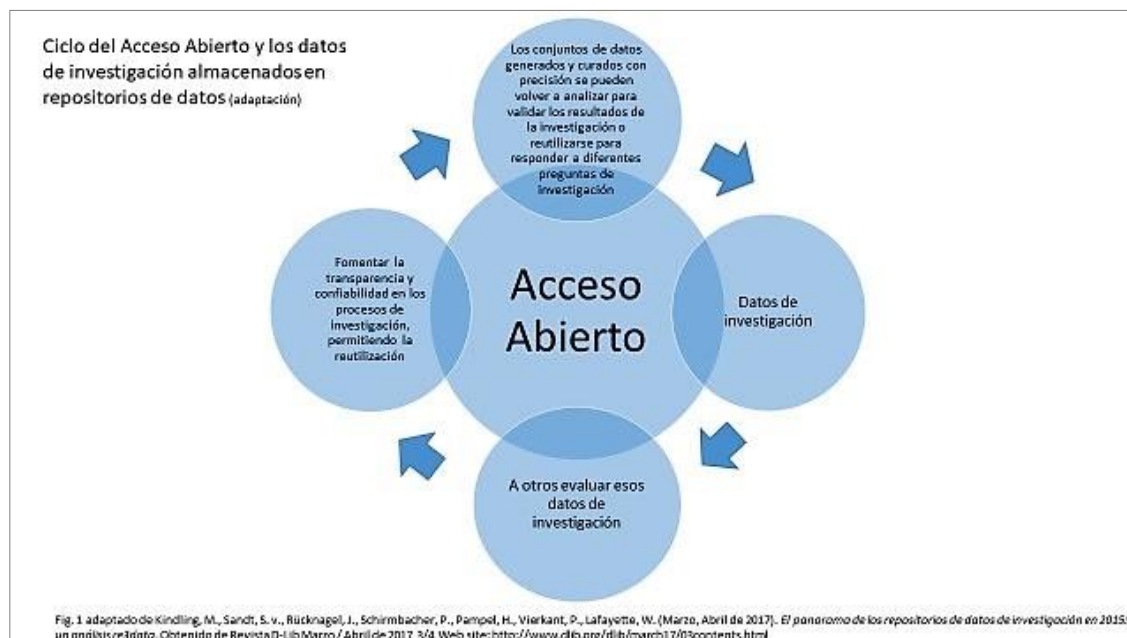


Figura 3-13: Ciclo de AA y los datos de investigación almacenados en repositorios de datos
Adaptado de Kindling, M. y otros. "El panorama de los repositorios de investigación en 2015: un análisis re3data". Obtenido de Revista D-Lib (marzo-abril 2017).

Como se ha señalado, re3data.org es reconocido hoy mundialmente por registrar repositorios de datos de investigación. Fue lanzado en 2013, permitiendo recopilar metadatos de los repositorios especializados en almacenar datos procedentes de investigación (Ministerio de Asuntos Económicos y Transformación Digital. Gobierno de España, 2017).

Se desarrolló a partir de un proyecto conjunto de varias organizaciones alemanas, mencionadas anteriormente, y fue financiado por la DFG. El lanzamiento oficial se produjo en mayo de 2013 y, posteriormente, se integró el catálogo Databib¹¹³ para evitar la duplicación y confusión por la

¹⁰⁹ Copernicus Publications. Es un editor de literatura científica con sede en Gotinga (Alemania), que fue fundado en 1994, publicando en la actualidad 28 revistas científicas de acceso abierto revisadas por pares y otras publicaciones en nombre de la Unión Europea de Geociencias. Sitio web: <https://www.copernicus.org/>. Consultado 210620.

¹¹⁰ PeerJ. Mega revista científica revisada por pares en AA para ciencias biológicas y médicas. Sitio web: <https://peerj.com/>. Consultado 210620.

¹¹¹ PLOS ONE. Revista científica en Acceso Abierto y publicada desde 2012 con este nombre por Public Library of Science para temas de investigación básica en cualquier materia relacionada con la ciencia y la medicina. Sitio web: <https://journals.plos.org/plosone/>. Consultado 210620.

¹¹² Nature's Scientific Data. Revista científica de Acceso Abierto y revisada por pares, publicada por Nature Publishing Group desde 2014. Sitio web: <https://www.nature.com/sdata/>. Consultado 210620.

¹¹³ DataBib. Sitio web: <https://researchlibrarian.net/2012/12/19/databib-databib-catalogo-registro-y-directorio-de-repositorios-de-datos-de-investigacion/>

existencia de dos registros similares paralelos. El proyecto de unificación estuvo auspiciado por DataCite que, como se señaló, busca mejorar la calidad de las citas de datos. Además, re3data.org colabora con otros proyectos de CA como es BioSharing, orientado sólo a Ciencias de la Vida y hoy conocido como FAIRsharing¹¹⁴, que ofrece estándares de metadatos y datos interrelacionados con bases de datos/repositorios y políticas de datos de todas las disciplinas u OpenAIRE¹¹⁵.

Por otra parte, hay referentes que proclaman que la promoción de la CA, la cultura del intercambio, la reutilización de información y el AA se encuentra en los cimientos del proyecto re3data.org (Ministerio de Asuntos Económicos y Transformación Digital. Gobierno de España, 2017).

Al abordar las tendencias relacionadas con la implementación de tecnologías (como son los repositorios de datos de investigación) surge DataCite. Pero, ¿qué es esta organización vinculada, en la actualidad, al principal directorio de repositorios de datos en Europa?

DataCite es una comunidad de bibliotecas, instituciones de investigación y centros de datos que albergan repositorios. Se unió a re3data en 2016.

Según su criterio, los depósitos de datos son importantes porque proporcionan servicios para datos de investigación, los cuales incluyen no sólo la mantención y la preservación de los datos, sino también hacer que éstos sean permanentemente accesibles y citables.

La principal actividad de DataCite es proporcionar herramientas para DOI y el registro de metadatos. Los DOI están registrados para contenido alojado en repositorios (en algunos casos, en revistas u otros contenedores para alojar contenido). Por lo tanto, existe una relación importante y natural entre el contenido (ya sea un conjunto de datos, un documento de texto o una muestra, por ejemplo) y el repositorio¹¹⁶ (Cousijn H. D., 2020).

Además del directorio re3data.org, desde 2012 existe otro portal concebido en Europa y que corresponde al Portal de Datos Abiertos¹¹⁷ de la CE, el cual proporciona datos geográficos, geopolíticos, financieros, medioambientales, estadísticos, gubernamentales y de investigación científica, entre otros, producidos por los países integrantes y organismos internacionales adheridos, los cuales se pueden usar y reutilizar -si así se desea- con fines no comerciales y también comerciales por parte de los ciudadanos, quienes pueden acceder a éstos y aprovechar su potencial económico. Este portal también aspira a que los datos depositados en los repositorios permitan la transparencia de la información proporcionada por los organismos participantes (EU. European Union, 2012).

¹¹⁴ Fairsharing Oxford. Sitio web: <https://fairsharing.org/> y <https://fairsharing.org/communities#team>

¹¹⁵ OpenAire para Ciencia Abierta en Europa. Sitio web: <https://www.openaire.eu/>

¹¹⁶ Consultado 210620.

¹¹⁷ Portal de Datos Abiertos de la Comunidad Europea. Sitio web: <https://data.europa.eu/euodp/es/data/>

Muchos países de Europa disponen de repositorios de datos de investigación regidos fuertemente por organizaciones integradas por potencias en las que existe una fuerte regularización y orientación a conversar temas que afectan el desarrollo de las diversas sociedades europeas.

Para el caso del Portal de Datos Abiertos, se ha tendido a facilitar el acceso e intercambio de los datos ahí alojados. De esta manera, es posible encontrar información clasificada en trece categorías creadas a partir de miles de conjuntos de datos (datasets) disponibles en variados formatos, tales como el Tab-Separated Values (TSV), que tiene la ventaja de visualizarse por cualquier software para hojas de cálculo, editores y procesadores de texto. Adicionalmente, los datos o set de datos más pesados (de mayor volumen) se comprimen en formato ZIP y el código abierto GC¹¹⁸.

Estas tendencias tecnológicas que se implementan y gestionan en la filosofía de AA, fuertemente respaldado y regulado en Europa, permiten la libre disponibilidad y uso de los datos, lo cual es un apreciado beneficio, propiciando así la transparencia y la confianza entre los diversos actores de la sociedad, lo cual repercute -a su vez- en el desarrollo social y económico de este continente.

En Europa existen sólidas redes regionales y nacionales de defensores del AA que representan bibliotecas y algunas comunidades de disciplina de investigación. Más de la mitad de los países europeos ya han establecido infraestructuras de repositorio nacionales (Schmidt, 2020). Indudablemente, esto repercute en la implementación y gestión de sus repositorios de datos de investigación abiertos.

Este desarrollo de infraestructuras tecnológicas fue impulsado desde OpenAIRE, creado en 2009 bajo el alero de la CE, una de las siete instituciones de la UE. A través de una serie de fases de proyectos financiados por la Comisión (desde los proyectos DRIVER para vincular la infraestructura de repositorios de Europa, hasta el primer proyecto OpenAIRE, destinado a ayudar a la CE a implementar su piloto inicial de AA) se ha consolidado la misión¹¹⁹ de OpenAIRE respecto de implementar políticas de AA a través de repositorios.

En la actualidad, OpenAIRE es mucho más que una robusta infraestructura tecnológica y de servicios dirigida a colaborar, acelerar y cuantificar la correcta implementación de las políticas europeas de AA a publicaciones científicas y datos de investigación.

Las dos primeras fases de OpenAIRE, denominadas DRIVER, se localizaron en Conductor I y II (2006 a 2009), siendo la primera iniciativa de la UE para explorar el desarrollo de una tecnología distribuida que permitiera una mayor interoperabilidad de los datos y que estableciera, además, una infraestructura de repositorios digitales. DRIVER creó una red de repositorios digitales en diez países

¹¹⁸ Gestión de Configuración (GC). A través de ella se establece un conjunto de tareas necesarias que deben implementarse en la organización para garantizar la calidad del software.

¹¹⁹ Cambiar la comunicación académica hacia la apertura y la transparencia, facilitando formas innovadoras de comunicar y monitorear la investigación.

Europeos, permitiendo una infraestructura de repositorio digital de calidad de producción que condujo, finalmente, al establecimiento de COAR¹²⁰.

Luego, se implementaron las fases estantes como OpenAIRE (2008-2011), donde la CE proporcionó los medios para promover y realizar la adopción generalizada de la política de AA; OpenAIREPlus (2012-2015), con el diferenciador plus, debido a que se logró ampliar las redes de repositorios de publicaciones para colaborar con proveedores de datos de áreas científicas específicas del dominio que apoyan la gestión y la interconexión entre datos de investigación, extendiéndose además a treinta y un países; OpenAIRE2020 (2015-2018), fase en la que el programa Horizonte 2020 extendió el piloto de FP7¹²¹ de AA a un mandato amplio para abarcar todos los proyectos, instituyendo un nuevo piloto para datos abiertos (cabe hacer notar que OpenAIRE2020 implementó y monitoreó estas actividades, promoviendo no sólo una beca abierta y mejorando la capacidad de descubrimiento y reutilización de las publicaciones y datos de investigación de H2020, sino también ampliando su cobertura a treinta y tres países); OpenAIREConnect (2015-2018), orientado a construir puentes tecnológicos y sociales para ofrecer servicios de intercambio uniforme de artefactos de investigación (literatura, datos y métodos), con vínculos semánticos entre comunidades de investigación y proveedores de contenido en comunicación científica; y, por último, OpenAIREAdvance (2018-2020), que se concibió para respaldar los mandatos de AA/datos abiertos en Europa, manteniendo su infraestructura humana y tecnológica, que consolida su conjunto de servicios orientados a impulsar CA¹²² en Europa (OpenAIRE, 2018).

3.11. Repositorios de datos de investigación en Estados Unidos

Actualmente, el directorio de re3data.org registra más de mil repositorios de datos de investigación en los Estados Unidos, superando en gran medida a los restantes países. Además de las iniciativas articuladas por las agencias federales e instituciones de investigación, las editoriales tienen aquí un activo rol en relación a establecer criterios para la selección de repositorios de datos de investigación, lo cual marca una diferencia en relación a Europa.

¹²⁰ Confederation of Open Access Repositories (COAR o Confederación de Repositorios de Acceso Abierto, en español). Es una asociación internacional con más de 150 integrantes y socios de todo el mundo que representan a bibliotecas, universidades e instituciones de investigación.

¹²¹ Séptimo programa marco de investigación y desarrollo. El programa de financiación del 7PM se desarrolló entre 2007 y 2013.

¹²² “Es un cambio radical en la forma de hacer investigación y difundir los resultados, que combina un gran desarrollo tecnológico y un cambio cultural hacia la apertura y la colaboración, promoviendo una ciencia más eficiente, transparente y accesible. Es lógico en el mundo digital, pero no es fácil”. Eva Méndez, Presidenta de la Plataforma de Política de Ciencia Abierta de la Comisión Europea. La ciencia abierta abarca todo el ciclo de producción del conocimiento científico, considerando la concepción, recogida de datos, procesamiento, publicación y distribución, al igual que la reutilización y evaluación de los resultados. Más datos en: <https://www.agenciasinc.es/Reportajes/Movimiento-open-science-un-nuevo-plan-para-liberar-a-la-ciencia-europea>

Editoriales y patrocinadores como PLOS ONE¹²³ y Nature, entre otras, han adoptado recomendaciones orientadas a que los investigadores compartan sus descubrimientos y hallazgos en repositorios que cumplan con una serie de criterios en sintonía, conjuntamente, con los Principios FAIR. De este modo, se logra optimizar la visibilidad de los datos y metadatos obtenidos por los investigadores, quienes deben acoger los estándares requeridos por su área de investigación.

Dentro de sus criterios, PLOS ONE establece que la nómina de repositorios por ellos sugerida no impone restricciones geográficas o de afiliación al envío de conjuntos de datos por parte de investigadores. En cuanto a los identificadores persistentes, los repositorios deben disponer de un Identificador Persistente Estable (Persistent Identifier, PID) para cada conjunto de los datos en la publicación, al igual que un Identificador de Objeto Digital (DOI, en inglés) o un número de acceso. Por otra parte, todos los datos deben ser publicados bajo las licencias CC-O o CC-BY (o equivalentes menos restrictivas). Los conjuntos de datos deben estar disponibles sin costo y sin requisitos de registro que pudiesen restringir el acceso. También deben disponer de un plan de gestión de datos, que incluya la preservación a largo plazo. Finalmente, deben ser repositorios reconocidos en la comunidad por parte de los investigadores y, asimismo, deben tener un acceso a FAIRsharing.org¹²⁴ (PLOS ONE Editors, 2018).

En cuanto a Nature, estipula también una serie de criterios y clasifica a los repositorios de datos de investigación en generalistas y disciplinares (SpringerNature, 2020). Este editor declara no alojar datos. Sin embargo, recomienda a los investigadores seleccionar opciones públicas que consideren aspectos tales como acceso, preservación y estabilidad. Respecto de la elección, también sugiere a los investigadores prestar atención sobre la existencia de políticas en algunos repositorios de datos, puesto que sólo aceptan resultados como producto de financiamientos específicos, advirtiéndolo incluso que pueden existir cobros por alojamiento de datos. En consecuencia, este editor sugiere revisar cuidadosamente las políticas de estos recursos, así como la nómina de repositorios de datos recomendados, los cuales deben estar disponibles para su uso bajo la licencia CC-BY, e insta a adoptar las sugerencias de FAIRsharing.

Algunos repositorios denominados generalistas -y recomendados por Nature- son presentados a continuación.

¹²³ Revista científica de Acceso Abierto que se fundó en 2006 como PLoS ONE. Su factor de impacto se estima de 3,234. Sitio web: <https://journals.plos.org/plosone/>

¹²⁴ FAIRsharing.org es un sitio que se construye con los aportes de investigadores que propician estándares, repositorios y políticas para almacenar datos en Acceso Abierto. Sitio web: <https://fairsharing.org/>

Tabla 3-3: Ejemplos de repositorios recomendados por Nature

Repository Name	Information on fees/costs	Size limits	Integrated with <i>Scientific Data's</i> manuscript submission system	Re3data / FAIRsharing entry
Dryad Digital Repository	\$120 USD for first 20 GB, and \$50 USD for each additional 10 GB	None stated	Yes ✓	view FAIRsharing entry
figshare	100 GB free per <i>Scientific Data</i> manuscript. Additional fees apply for larger datasets	1 TB per dataset	Yes ✓ - To qualify for the 100 GB of free storage, data must be uploaded to figshare via our submission system. Download instructions.	view FAIRsharing entry
Harvard Dataverse	Contact repository for datasets over 1 TB	2.5 GB per file, 10 GB per dataset	No	view re3data entry
Open Science Framework	Free of charge	5 GB per file, multiple files can be uploaded	No	view FAIRsharing entry
Zenodo	Donations towards sustainability encouraged	50 GB per dataset	No	view re3data entry
Mendeley Data	Contact repository for datasets over 10 GB	10 GB per dataset	No	view FAIRsharing entry

3.12. Repositorios de datos de investigación en América Latina - LA Referencia

Actualmente, LA Referencia está conformado por diez nodos: Argentina, Brasil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, México, Perú y Uruguay, considerándose una plataforma relevante que aspira a contener la producción en AA de América Latina.

LA Referencia dispone de prestaciones tecnológicas para los nodos regionales, entre los cuales cabe hacer notar:

➤ Desarrollo tecnológico para cosechadores nodales:

Este desarrollo tecnológico se originó en 2013 y, desde 2015, esta plataforma de software libre (GPL 3)¹²⁵ se implementó para recolectar información que, además de soportar el nodo central de la red regional (Brasil), funcione como servicio “agregador”, entendiendo por éste un software, plataforma web o aplicación que permita aunar en un solo espacio datos existentes en distintas plataformas

¹²⁵ General Public License (GPL 3). Correspondiente a la versión 3 (2007), es una licencia de derecho de autor ampliamente usada en el mundo del software libre y del código abierto, garantizando a los usuarios finales (personas, organizaciones, compañías) la libertad de usar, estudiar, compartir (copiar) y modificar el software. Su propósito es doble: por un lado, declarar que el software cubierto por esta licencia es libre y, por otro, protegerlo -mediante una práctica conocida como copyleft- de intentos de apropiación que restrinjan dichas libertades a nuevos usuarios cada vez que la obra es distribuida, modificada o ampliada. Sitio web: <https://www.gnu.org/licenses/gpl-3.0.html> Consultado 230620.

digitales. Esta herramienta está en el campo de las nuevas tecnologías de información y comunicaciones (ANOVO, 2019).

A partir de los nodos nacionales, se integran artículos científicos y tesis provenientes de más de un centenar de universidades e instituciones de investigación de los diez países que hoy conforman:

- 1.- Argentina (Ministerio de Ciencia, Tecnología e Innovación, MinCyT)¹²⁶
- 2.- Brasil (Instituto Brasileño en Información Ciencia y Tecnología, IBICT)¹²⁷
- 3.- Colombia (Red Nacional Académica de Tecnología Avanzada, RENATA)¹²⁸
- 4.- Costa Rica (Consejo Nacional de Rectores, CONARE)¹²⁹
- 5.- Chile (Comisión Nacional de Investigación, Ciencia y Tecnología, CONICYT, hoy Agencia Nacional de Investigación y Desarrollo, ANID)¹³⁰
- 6.- Ecuador (Corporación Ecuatoriana para el Desarrollo de la Investigación y la Academia, CEDIA, junto con Secretaría de Educación Superior, Ciencia, Tecnología e Innovación, SENECYT)¹³¹
- 7.- El Salvador (en instalación Consorcio de Bibliotecas Universitarias de El Salvador, CBUES, del Ministerio de Ciencia y Tecnología)¹³²
- 8.- Perú (Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica, CONCYTEC)¹³³
- 9.- Uruguay (Agencia Nacional de Investigación e Innovación, ANII)¹³⁴
- 10.- México (en evaluación con un repositorio versión Beta)¹³⁵

LA Referencia dispone también de tres componentes tecnológicos los cuales le permiten cumplir con el objetivo de la recolección digital de los nodos:

- **LRHarvester**¹³⁶: Chile cuenta con esta plataforma, la cual permite la cosecha en el repositorio. Permite también la transformación y validación de metadatos Open Archive Initiative-Protocol for Metadata Harvesting (OAI-PMH). Responsable de la recolección de fuentes OAI, validación de registros de metadatos y su transformación para cumplir con las pautas de calidad de la red. La plataforma está desarrollada en lenguaje Java 1.7, utilizando el framework Spring y el motor de base de datos PostgreSQL 9.X.

¹²⁶ Sitio web: <https://www.argentina.gob.ar/ciencia>

¹²⁷ Sitio web: <http://www.ibict.br/>

¹²⁸ Sitio web: <https://www.renata.edu.co/>

¹²⁹ Sitio web: <https://www.conare.ac.cr/>

¹³⁰ Sitio web:

<https://www.anid.cl/2020/01/14/agencia-nacional-de-investigacion-y-desarrollo-es-la-sucesora-de-conicyt/>

¹³¹ Sitio web: <https://www.educacionsuperior.gob.ec/senescyt-y-cedia-implementaran-plataforma-web-que-recogera-la-oferta-tecnologica-de-las-instituciones-de-educacion-superior-e-institutos-publicos-de-investigacion-y-la-demanda-de-los-sectores-producti/>

¹³² Sitio web: <http://www.lareferencia.info/es/nodos/el-salvador>

¹³³ Sitio web: <https://portal.concytec.gob.pe/>

¹³⁴ Sitio web: <https://www.anii.org.uy/>

¹³⁵ Sitio web: <http://www.lareferencia.info/es/nodos/mexico>

¹³⁶ Sitio web: <http://www.lareferencia.info/es/recursos/capacitacion/91-lautaro-matas>

- **LRProvider**¹³⁷: Chile cuenta con este componente responsable de publicar metadatos utilizando el protocolo OAI-PMH, actualmente derivado del software DSpace-XOAI, que ofrece alta flexibilidad y escalabilidad para la interoperabilidad con otros agregadores internacionales de metadatos.
- **Search Engine/Service Portal**: Se trata de un portal web de publicación de metadatos agregado que proporciona una interfaz de búsqueda amigable y adaptativa para dispositivos móviles. También proporciona estadísticas sobre la calidad de los metadatos y la evolución de los cultivos. Este componente se basa en el software libre VuFind¹³⁸.

En relación a los principios de diseño definidos en 2014-2015, LA Referencia declara que éstos continúan siendo válidos:

- Transferible: la versión del nodo central debe ser la misma que la de los nodos nacionales.
- Responsivo: acceso por dispositivos móviles (bootstrap).
- Multilingüe: soporte en inglés y portugués.
- Orientado a los negocios: enfoque en validaciones y transformaciones de metadatos en OAI.
- Soporte para pautas de OpenAIRE y flexibilidad para actualizar nuevas pautas.
- Escalable: debe cosechar cientos de repositorios y más de un millón de registros en menos de 24 horas.
- Administrable y amigable: herramientas de diagnóstico, validación e informes para administradores, además de la posibilidad de manejo por especialistas en información y no sólo por técnicos informáticos.
- OpenSource: desarrollo y uso de componentes y soluciones (por ejemplo, VuFind para el Front-end o usar bases de datos conocidas como MySQL, PostGres y lenguaje Java).
- Estadísticas: desarrollo de nuevos módulos.
- Economical: basado en distribuciones gratuitas de Linux como Ubuntu y un nodo nacional “promedio”, una máquina de dos núcleos con 8 GB de RAM, 500 GB de espacio¹³⁹ (LA Referencia. Red de Repositorios de Acceso Abierto a la Ciencia, 2020).

LA Referencia aborda temas de interoperabilidad, a través de su documento “Metadatos y políticas de recolección”¹⁴⁰, lanzado en 2015 y en el cual se establece una serie de pautas de

¹³⁷ Sitio web: <https://github.com/DSpace/xoai>

¹³⁸ VuFind. Desarrollado con módulos específicos, es un portal de recursos de biblioteca diseñado y desarrollado por bibliotecas y para bibliotecas. El objetivo de VuFind es permitir a sus usuarios buscar y navegar por todos los recursos de su biblioteca, reemplazando el OPAC tradicional para incluir: registros de catálogos, diarios almacenados en caché local, elementos de la biblioteca digital, repositorio institucional, bibliografía institucional y otras colecciones y recursos de la biblioteca. Sitio web: <https://vufind.org/vufind/>

¹³⁹ Consultado 230620.

¹⁴⁰ Documento que contiene recomendaciones de implementación sobre la base de acuerdos internacionales de interoperabilidad, orientados respecto de cómo poner los recursos científicos digitales utilizando el protocolo OAI-PMH y los metadatos Dublin Core. Proviene de DRIVER, absorbida por lo que hoy es OpenAIRE, ésta última encargada de las directrices tecnológicas y de los Current Research Information Systems (CRIS), una infraestructura que sustentará el programa Horizonte 2020 en cuanto al AA de la Información. Sitio web:

interoperabilidad cuyo cumplimiento debe ser garantizado por los nodos nacionales, recomendando al mismo tiempo su adopción por los repositorios que conforman la red. Las directrices acordadas a nivel regional se basan en las Directrices Driver 2.0 y las Directrices OpenAIRE.

En noviembre de 2018 se publicó la nueva versión de las Directrices 4.0, en el que colaboró el equipo técnico de LA Referencia, siendo aprobado por la Junta Directiva de la misma a fines de 2018. Su implementación será gradual desde 2020.

3.13. Repositorios de datos de investigación en Chile

Cabe destacar que, para el caso de Chile, se encontraron registrados en re3data.org dos repositorios de datos.

Uno de ellos se encuentra en la Región Metropolitana y corresponde a un Repositorio de Datos de Investigación de la Universidad de Chile¹⁴¹, el cual se encuentra conformado por catorce facultades que han aportado con 13 conjuntos de datos y 51 archivos, desde 2019 a octubre de 2020. Este repositorio de datos se implementó en Dataverse con licencia Creative Commons y con las tres clasificaciones para el AA de sus datos: restringido, abierto y con embargo. En cuanto a las normas, este repositorio de datos cuenta con el esquema estándar de metadatos de la Digital Centre Curation (DCC)¹⁴², de Reino Unido, y de Dublin Core™ Metadata Initiative¹⁴³.

El segundo repositorio de datos corresponde a una colaboración entre la Universidad Austral y el Queensland Brain Institute, de la Universidad de Queensland (Australia). Este repositorio, llamado e-Fish¹⁴⁴, fue lanzado en 2015. En cuanto a la clasificación de sus datos, estos se encuentran en AA y restringido. También dispone de una licencia Creative Commons de dominio público. No declara el nombre del software del repositorio en re3data.org; sin embargo, cuenta con servidores dedicados en la Universidad de Queensland y declara contar con tecnología NGS¹⁴⁵ (Secuenciación de Nueva Generación, en español), tecnología que emerge a mediados de la década del 2000 y que se ha adoptado en algunos repositorios de datos, preferentemente especializados, con el propósito de realizar análisis de datos. Con esto se logra optimizar el procesamiento de la información y

<http://www.lareferencia.info/es/recursos/directrices-metadatos/4-metadatos-y-politicas-de-cosecha-interoperables-para-los-nodos-nacionales> Consultado el 24 de junio de 2020.

¹⁴¹ Repositorio de Datos de Investigación de la Universidad de Chile. Sitio web: <https://datos.uchile.cl/>

¹⁴² Digital Centre Curtion (DCC). Sitio web: <https://www.dcc.ac.uk/resources/metadata-standards/repository-developed-metadata-schemas>

¹⁴³ Dublin Core™ Metadata Initiative. Sitio web:

<https://www.dublincore.org/specifications/dublin-core/dcmi-terms/dcmitype/Dataset/>

¹⁴⁴ e-Fish. Sitio web: <http://www.molecularfisherieslaboratory.com.au/efish/>

¹⁴⁵ Next-Generation Sequencing (NGS). Término con el que se hace referencia a las tecnologías diseñadas para analizar gran cantidad de ADN de forma masiva y paralela. Sitio web: <https://www.elsevier.es/es-revista-agnostico-prenatal-327-articulo-tecnologias-secuenciacion-nueva-generacion-agnostico-S2173412712000273>

repercute, por otro lado, en una reducción de costos de mantención de esta tecnología para la institución. Este repositorio declara albergar más de 120 millones de conjuntos de datos en NGS.

4. Alcances y limitaciones

4.1. Alcances

El alcance de este trabajo está orientado a lograr los objetivos propuestos, así como permitir dar respuesta a la siguiente pregunta de investigación: ¿Qué beneficios y desafíos implica para las bibliotecas la implementación y gestión de un repositorio de datos de investigación?

La metodología empleada se sustentó en una exploración y posterior descripción, la cual permitió dar respuesta a la interrogante inicial del trabajo. Para alcanzar lo anterior, la metodología incluyó una revisión bibliográfica del tema, íntegramente en fuentes de información en Acceso Abierto (AA) desde 2017 a la fecha y con variadas estrategias de búsqueda, incluyendo palabras clave asociadas a la temática, tales como CA, AA, directorios de repositorios de datos de investigación, repositorios de próxima generación, estándares de repositorios de datos, metadatos de repositorios de datos, softwares de repositorios de datos de investigación, políticas de AA para datos de investigación, identificadores persistentes, identificador de objeto digital, licencias, OpenAIRE, entre otras. La cobertura idiomática fue en español y en inglés, mientras que la cobertura geográfica incluyó Europa, Estados Unidos, Latinoamérica y Chile.

Se revisó exhaustivamente la bibliografía recolectada, seleccionando sólo aquellas fuentes cuyo contenido se ajustase a la interrogante inicial planteada para este tema. Complementario a lo anterior, realicé un benchmarking de los sitios web especializados en la materia.

Por otro lado, tanto la exploración como la descripción se enfocaron en proporcionar una recopilación de información para bibliotecas o instituciones que requieran implementar, gestionar o aplicar mejoras en un repositorio de datos de investigación, encontrándose éste en una etapa inicial.

En dicho sentido, se proporciona una visión de las tendencias de desarrollo en esta materia, experimentado a nivel de continentes, de disciplinas y de aspectos de jurisprudencia, los cuales permiten conocer y apreciar las diferencias alcanzadas entre los países y su cercanía a la generación de políticas de AA y/o de datos de investigación. Al ser un factor que repercute, se sugiere que sea considerado al momento de implementar y gestionar un repositorio de datos de investigación abierto en bibliotecas.

Asimismo, para articular las respuestas planteadas en el presente trabajo, me apoyé también en entrevistas online, como fue el caso con el académico e investigador Agustín Fernández Eguiarte, Director del Repositorio Institucional del Centro de la Atmósfera de la Universidad Nacional Autónoma de México (UNAM), con quien dialogué en extenso luego de concertar de común acuerdo una reunión virtual a través de la plataforma Zoom.

Por otro tanto, con Washington Segundo, Doctor en Informática y Coordinador del Laboratorio de Metodología de Tratamiento y Difusión de la Información (COLAB) en el Instituto Brasileño de Información en Ciencia y Tecnología (IBICIT/ MCTIC), al igual que con Andrew McKenna-Foster, Especialista en Producto de Figshare, intercambié algunos correos electrónicos que orientaron y/o respaldaron algunos aspectos específicos de mi investigación, accediendo a contenidos que aportaron en valor y calidad.

En tanto, el análisis de los dos casos incluidos en el presente trabajo (uno de ellos correspondiente a la UNAM y el otro a Figshare, que se ilustra con un ejemplo específico y exitoso, como es la Universidad Estatal de Iowa) también me permitió compenetrarme en profundidad de aspectos prácticos y hasta de recursos humanos que conlleva la implementación de un repositorio de datos de investigación. Gracias a este análisis, me fue posible advertir una serie de eventos asociados no sólo a la elección de un repositorio, sino también a su implementación y desarrollo, que puede convertirse en una solución eficiente y exitosa cuando todos los agentes de la cadena se encuentran alineados en un mismo sentido.

Con el propósito de entregar una mirada lo más global posible, también planifiqué una encuesta - hasta ahora inédita en su tipo-, por medio de la cual quise recoger las experiencias de diversas instituciones prestigiosas y consolidadas, tanto en Chile como en América Latina, Estados Unidos, Canadá, Europa, Reino Unido y Australia, en la implementación y desarrollo de repositorios de datos de investigación. A través de dicho instrumento, quise conocer -además- sus experiencias particulares en dicho proceso, instándolos a compartir sus buenas prácticas con el propósito de que sirvan de estímulo y de apoyo a otras instituciones que recién se suman (o que piensan hacerlo) a esta fascinante comunidad virtual de información accesible para todos. A través de 15 preguntas que conformaron la encuesta -de la cual se hicieron dos envíos (entre el 13 y 15 de octubre y el 3 noviembre) a un total de 333 instituciones- obtuve información relevante y de primera fuente sobre softwares, financiamiento, recursos humanos comprometidos, legislación, antigüedad y tipología del repositorio, lo cual hace de ésta -a la fecha- una investigación vanguardista en su enfoque.

Todos los aspectos descritos anteriormente se analizarán en detalle más adelante, en los capítulos de Metodología, Aplicación a Casos de Estudios y Resultados y Análisis.

4.2. Limitaciones

En relación a las limitaciones, la concepción de este trabajo se relaciona con la CA y el AA, por lo que el contenido aborda soluciones tecnológicas o presentaciones de casos en esta condición.

Por otra parte, la recopilación de información se enfoca principalmente en proporcionar recomendaciones y/o buenas prácticas, las cuales permitan la implementación y gestión de un repositorio de datos. Es una tesina que pretende ser una publicación de utilidad para aquellas

instituciones que requieran emprender un proyecto tal, como por ejemplo la creación o desarrollo de un repositorio de datos de investigación.

Este trabajo se centra fundamentalmente en los datos de investigación y en los repositorios, abordando mínimamente las publicaciones académicas y/o de investigación en AA.

En relación al ciclo de vida de los datos de investigación, el trabajo enfatiza los repositorios de datos, los cuales son un componente en relación a Research Data Management (RDM) o plan de gestión de datos para la investigación.

Esta publicación sugiere que algunos puntos puedan desarrollarse en mayor profundidad, porque también impactan en la implementación y en la gestión de un repositorio de datos:

- Gestión de recursos humanos y la estructura de personal requerida para administrar un repositorio de datos de investigación.
- Un análisis completo de costos directos e indirectos de esta implementación.
- La gestión de cambio en la organización afectada por la implementación de un proyecto de esta envergadura.

5. Objetivos

5.1. General

Recopilar y entregar pautas y recomendaciones para la implementación, gestión y mejora continua de un repositorio de datos de investigación.

5.2. Específicos

- Analizar una muestra relevante de repositorios de datos existentes en el mundo para conocer y comprender sus principales características.
- Presentar el estado de la CA y el AA, al igual que sus comportamientos en distintos continentes.
- Identificar los recursos y las buenas prácticas utilizadas en los repositorios de datos de investigación.
- Presentar la situación en Chile respecto de la implementación de repositorios de datos de investigación.

6. Situación actual

Así como hace casi 30 años varios países de Europa se unieron para conformar la Unión Europea (UE), cuyo objetivo fue unificar políticas de corte comercial y económico, destrabando -por ejemplo- el desplazamiento de las personas a través de las fronteras, y dando vida a una moneda común (el Euro), desde comienzos del siglo XXI en este mismo continente se está produciendo un fenómeno similar, aunque esta vez en términos de las ciencias. Motivados por la necesidad de hacer del conocimiento científico un bien de uso público y, por ende, accesible a cualquier persona -además de cumplir con estándares de calidad, tecnológicos y de gestión, como se verá más adelante-, un grupo de países se alió para dar vida a una infraestructura tecnológica y de servicios conocida como OpenAIRE, la cual ha ido evolucionando desde su puesta en funcionamiento en 2009. Desde entonces, los 34 países que actualmente la integran han dado pasos decisivos en su regulación y funcionamiento, generando políticas y estándares de calidad con respecto a lo que se entiende por Ciencia Abierta (CA), Acceso Abierto (AA) y repositorios -dentro de éstos, los repositorios de datos de investigación-, y respecto de la infraestructura tecnológica definida para albergar los resultados generados en esta nueva forma de ciencia compartida, incentivando una serie de indicaciones concretas y recomendaciones para otros países que, eventualmente, quieran sumarse dentro o fuera del llamado Viejo Continente (OpenAIRE, s.f.).

Algo similar acontece en Estados Unidos, Australia y Canadá, países líderes en conocimiento científico y en AA, los cuales han avanzado prácticamente en sintonía con los europeos respecto de este tema. Hoy, el grueso de las instituciones académicas, al igual que muchas organizaciones que producen y almacenan datos científicos, disponen de repositorios normados según las directrices planteadas desde OpenAIRE, alineándose con los protocolos indispensables para facilitar el avance del conocimiento científico por medio de la calidad y de la cooperación.

Respecto de Latinoamérica, hay una desnivelación acentuada en comparación a Europa, Estados Unidos y Australia, tanto entre los países que conforman esta zona geopolítica como dentro de cada país por sí mismo. De hecho, en la actualidad, sólo 10 integran LA Referencia, una red de repositorios de AA a la ciencia que también ha debido amoldarse a las directrices emanadas desde OpenAire y que se encuentra abocada a reclutar más integrantes, con el fin de cosechar publicaciones en una mayor cantidad de nodos, al tiempo que no sólo regula, sino también incentiva, los estándares que deben adscribir las instituciones y organismos para ser aceptados en la comunidad científica internacional (Amaro, 2020). En Chile, que forma parte de LA Referencia, se realizó recientemente una encuesta sobre política de AA (Desarrollo, 2020), arrojando interesantes resultados sobre los cuales me referiré más adelante, en este mismo capítulo.

Con ventaja frente a Latinoamérica se encuentra Brasil, donde también han entendido la necesidad de estar en sintonía con el mundo desarrollado, a pesar de que aquí ya se había comenzado a dar pasos firmes en el tema de repositorios de datos de investigación, promoviendo la CA, el AA y los datos como una consecuencia natural de las investigaciones científicas desarrolladas con recursos

públicos. Hoy, este país se encuentra en una fase de descentralización que vino a reemplazar el modelo centralizado. La idea es llegar a un modelo distribuido -como se explicará al abordar la situación actual de Brasil- donde cualquier institución, independientemente de su envergadura y localización geográfica dentro del país, pueda generar investigaciones con datos de calidad, en línea con los principios y estándares de calidad internacionales, para luego ser cosechadas por otras infraestructuras globales como LA Referencia, en Latinoamérica, y Open Aire, en Europa.

6.1. Unión Europea

Sin duda, OpenAIRE (Access Infrastructure Research for Europe o Infraestructura de Acceso Abierto para la Investigación en Europa, en español) ha sido un proyecto clave en la difusión y consolidación de CA, AA y datos de investigación, liderando por lejos en estas materias y con un objetivo muy claro, cual es promover la implementación generalizada de esta práctica para “cambiar la comunicación académica hacia la apertura y la transparencia, y facilitar formas innovadoras de comunicar y monitorear la investigación” (OpenAIRE, s.f.). La idea es que el conocimiento científico validado que generan los investigadores sirva a las personas y a otros estamentos sociales para construir un entorno de mayor calidad.

En la actualidad, los siguientes 34 países conforman las llamadas Mesas Nacionales de Acceso Abierto (NOAD, en inglés): Alemania, Austria, Bélgica, Bulgaria, Croacia, Chipre, Dinamarca, Eslovaquia, Eslovenia, España, Estonia, Finlandia, Francia, Grecia, Hungría, Irlanda, Islandia, Israel, Italia, Letonia, Lituania, Luxemburgo, Malta, Noruega, Países Bajos, Polonia, Portugal, Reino Unido (en proceso de separación debido al Brexit), República Checa, Rumanía, Serbia, Suecia, Suiza y Turquía.

Dispone de un fondo de 360 mil euros para descubrir, apoyar e implementar ideas que fomenten la CA, alojando millones de registros de metadatos vinculados, los cuales han sido recopilados de 10 mil fuentes de datos de confianza académica. Lo anterior es posible gracias a la práctica de las llamadas “Las tres O” (OpenAIRE, s.f.):

- Open Innovation (OI o Innovación Abierta, en español): centrada en nuevos modelos de negocio y esquemas de financiamiento, promueve el acceso al conocimiento con el fin de reutilizarlo en la generación de nuevos conocimientos, productos o servicios, garantizando el flujo libre de conocimientos, datos y nuevas tecnologías para pasar de la investigación a la innovación.
- Open Science (OS o CA, en español): se refiere al proceso y a la práctica de la investigación digitalizada, considerando el ciclo completo de una investigación, desde que se formula una idea, pasando por la selección del material de apoyo (literatura, datos), hasta llegar a la publicación, archivo y reutilización de los resultados de la investigación a que dio lugar dicha idea, donde se incluyen los metadatos y los datos utilizados.

- Open to the World (OW o Abierto al Mundo, en español): alude a que el conocimiento generado en las investigaciones, al igual que los datos asociados a éstas, son de utilidad para los desafíos de toda índole no sólo dentro de Europa, sino también en el resto del mundo, con el propósito de generar soluciones útiles y productivas.

En su evolución para convertirse en la infraestructura tecnológica que hoy es, OpenAIRE ha pasado por seis períodos, mencionados anteriormente, pero que me pareció importante destacar también en este capítulo (OpenAIRE, s.f.):

- DRIVER I (2006-2007) y DRIVER II (2007-2009): primera iniciativa de la Comisión Europea (CE) para propiciar el desarrollo de una tecnología destinada a mejorar la interoperabilidad de los datos por medio de una infraestructura paneuropea y comunitaria de repositorios digitales. La primera red se expandió a 10 países y se tradujo, además, en el establecimiento de la Confederación de Repositorios de Acceso Abierto (COAR, en inglés)¹⁴⁶.
- OpenAIRE (2008-2011): nació para apoyar la implementación del AA en Europa, siendo financiado por el 7° Programa Marco de Investigación y Desarrollo (7PM) de la UE. Proporcionó los medios para promover y materializar la adopción generalizada de la política de AA, según las directrices del European Research Council (ERC o Consejo Científico Europeo, en español) para el AA y el programa piloto de AA lanzado por la CE.
- OpenAIREplus (2012-2015): se trata de la segunda generación de infraestructura de AA para la investigación en Europa, expandiendo las redes de repositorios de publicaciones para colaborar con proveedores de datos de áreas científicas específicas de dominio, que respaldan la gestión y la interconexión entre datos científicos. La red se expandió a 31 países y comenzó a contribuir en la generación de estándares comunes, en protocolos para datos y en la interoperabilidad a nivel mundial.
- OpenAIRE2020 (2015-2018): implementó y monitoreó los proyectos de AA del 7PM desarrollados al alero del programa Horizonte 2020¹⁴⁷, que -a su vez- debe desarrollar un Plan de Gestión de Datos¹⁴⁸ que proporcione AA a los datos de investigación. Asimismo, se considera la antesala de Horizonte Europa¹⁴⁹, que debería desarrollarse entre 2021-2027.

¹⁴⁶ Confederation of Open Access Repositories (COAR) es una asociación internacional fundada en 2009 y que, en la actualidad, cuenta con 157 integrantes y socios de todo el mundo que representan a bibliotecas, universidades, instituciones de investigación, financiadores gubernamentales y otros. Reúne repositorios individuales y redes de repositorios para crear capacidad, alinear políticas y prácticas, y actuar como una voz global para la comunidad de repositorios. Sitio web: <https://www.coar-repositories.org/>

¹⁴⁷ Horizonte 2020 se considera el instrumento principal de la UE para financiar actividades de investigación y desarrollo en el período 2014-2020, privilegiando la participación individual o por medio de consorcios de centros de investigación, universidades, empresas e investigadores independientes, con un presupuesto de 80 mil millones de euros. Sitio web: <https://ec.europa.eu/programmes/horizon2020/en>

¹⁴⁸ El Plan de Gestión de Datos es un documento formal que especifica cómo se manejarán los datos de investigación durante y después de un proyecto de investigación. Identifica acciones y estrategias clave para garantizar que los datos de investigación sean de alta calidad, seguros, sostenibles y, cuando sea posible, accesibles y reutilizables. Cada vez más financiadores de investigación requieren un DMP como parte del proceso de propuesta de subvención, o después de que se haya aprobado el financiamiento. Sitio web: <https://www.openaire.eu/how-to-comply-to-h2020-mandates-for-data>

¹⁴⁹ Programa Marco de Investigación e Innovación de la Unión Europea, dotado con un presupuesto de 100 mil millones de euros y cuyo objetivo es impulsar la capacidad de innovación, la competitividad y el empleo de Europa, en el período 2021-2027. Sitio web: <https://ec.europa.eu/>

- OpenAIRE-Connect (2015-2018): ofrece servicios para un intercambio uniforme de artefactos de investigación (literatura, datos y métodos) con vínculos semánticos entre ellos, a través de comunidades de investigación y proveedores de contenido en comunicación científica.
- OpenAIRE Advance (2018-2020): continúa con la misión de OpenAIRE en cuanto a apoyar los mandatos de AA y CA en Europa, manteniendo su infraestructura humana y tecnológica.

Sobre los archivos de datos, OpenAIRE recomienda: Identificador (M), Creador (M), Título (M), Editor (M), Año de publicación (M), Asunto (R), Contribuyente (MA/O), Fecha (M), Idioma (R), ResourceType (R), Alternate Identifier (O), Related Identifier (MA), Tamaño (O), Formato (O), Versión (O), Derechos (MA), Descripción (MA) y Geolocalización (O) (OpenAIRE, 2015).

Respecto de la selección de un repositorio de datos, OpenAIRE sostiene que “la elección puede ser sencilla si tiene una instalación de gestión de datos establecida en su dominio o institución, o - incluso- dentro de su grupo o departamento de investigación. Cuando existan estándares o normas de preservación de datos en su disciplina, éstos deben seguirse. Su financiador de investigación puede recomendar un centro de datos o un archivo de autodepósito” (OpenAIRE, 2015).

Asimismo, insta a tomar en cuenta factores como la asignación de un Identificador Único y Persistente al conjunto de datos para asegurar que “los resultados de la investigación en repositorios dispares puedan vincularse a investigadores y subvenciones particulares”; que permita una página de destino para cada conjunto de datos, con metadatos que ayuden a encontrarlo, definirlo, relacionarlo con otras publicaciones y citarlo; que ayude a rastrear cómo se han utilizado los datos mediante estadísticas de acceso y descarga; que responda a las necesidades de la comunidad mediante una certificación como “repositorio de datos confiable”; que se adapte a las necesidades particulares de datos (formatos aceptados, acceso/copia/seguridad y sostenibilidad del servicio); que ofrezca términos y condiciones ajustados a la normativa legal; y que proporcione directrices sobre cómo citar los datos que se depositarán en el repositorio (OpenAIRE, 2015).

OpenAIRE considera adecuado que la institución ofrezca soporte de gestión de datos a los investigadores, lo cual se sustenta en la necesidad de registrar tanto los conjuntos de datos como el catálogo de datos de la institución para facilitar la visibilidad de la publicación. Se destaca la necesidad de depositar el conjunto de datos en un repositorio institucional no sólo para custodiarlo, sino también para tener acceso a datos estadísticos, asesorar sobre metadatos enriquecidos y seleccionar una licencia adecuada para usar los datos de una investigación (OpenAIRE, 2015).

Entre los repositorios de datos, se recomiendan Zenodo (que tiene la ventaja de ser gratuito) y re3data.org.

En cuanto a los Principios FAIR, OpenAIRE los promueve entusiastamente, ya que a través de su fomento se favorecería la CA, el AA y la calidad de los datos de investigación, llamamiento que parece tan obvio pero que, sin embargo, ha encontrado resistencia de parte de algunos

investigadores, quienes consideran que no hay más reconocimiento que ser validado por otro autor, aunque apenas en una cita. Por otra parte, esta aparente renuencia se relaciona más bien con las dificultades de software con que aquellos se encuentran, como evidenció una investigación realizada por PLOS ONE, que demostró que sólo 1/5 de las publicaciones comparte sus datos de investigación en repositorios de datos. “Los científicos no comparten datos por muchas razones. Aquellos que crean datos rara vez reciben crédito y, cuando lo hacen, el reconocimiento a menudo se limita a las citas. Hay escaso soporte disponible para curar datos. Estos temas abarcan todas las disciplinas, pero las conversaciones están desconectadas” (OpenAIRE, s.f.).

De allí iniciativas como las emprendidas por las revistas Springer Nature y Scientific Data que, recientemente, suscribieron la Declaración de Compromiso en las Ciencias de la Tierra, el Espacio y el Medio Ambiente. “Los editores insistirán en que los datos clave estén disponibles en repositorios que respalden los Principios FAIR. Estos cambios en la política y la práctica elevan los datos a valiosas contribuciones de investigación en lugar de archivos que se introducen como una ocurrencia tardía. Prometen abrir vías de descubrimiento científico y mejorar la replicabilidad” (COPDESS. Coalition for Publishing Data in the Earth and Space Sciences, s.f.).

Cabe recordar que los Principios FAIR se publicaron, formalmente, en la revista Scientific Data de Nature Publishing Group el 15 de marzo de 2016 (SciELO , 2016), en lo que se considera un acuerdo para uniformar el manejo y la administración de datos científicos, siendo éstos últimos la producción principal de una investigación científica ya publicada, de los que complementariamente también se promueve su reutilización para garantizar su validez, reproductibilidad y conducción a nuevos descubrimientos. Para alcanzar el objetivo descrito, los Principios FAIR establecen que los datos de investigación deben cumplir cualidades precisas y mensurables, como ser Encontrables (Findable), Accesibles (Accessible), Interoperables (Interoperable) y Reutilizables (Reusable).

Una actualización de los Principios FAIR se efectuó el pasado 30 de octubre, por parte de la CE y el Comité Ejecutivo de la European Open Science Cloud (EOSC o Nube Europea de Ciencia Abierta, en español). La publicación “Seis recomendaciones para la implementación de la práctica FAIR” surgió del trabajo en conjunto entre integrantes de FAIR y de la EOSC, actualizando normativas y evidenciando la relevancia de este tema (GOFAIR, 2020):

- Financiar la sensibilización, la formación, la educación y el apoyo específico de la comunidad.
- Financiar el desarrollo, adopción y mantenimiento de estándares, herramientas e infraestructura de la comunidad.
- Incentivar el desarrollo de la gobernanza comunitaria
- Traducir las directrices FAIR para otros objetos digitales.
- Premiar y reconocer las mejoras de la práctica FAIR.
- Desarrollar y monitorear políticas adecuadas para datos FAIR y objetos de investigación.

Actualmente, se considera que un cambio en la cultura de investigación en todas las disciplinas supone la ocurrencia de tres variables (Nature, 2019):

- Que tanto el repositorio de datos abiertos como la implementación de los Principios FAIR sean una prioridad para todos los estamentos involucrados en una publicación, considerando mandatos y orientaciones más estrictas para alinear dichas acciones. Al respecto, cabe mencionar iniciativas como el Informe 6 del Grupo de Expertos de la CE sobre datos FAIR, las guías de formación en línea y páginas web de autoayuda para aportar a los investigadores en cuanto a citar datos y software correctamente desarrolladas por el grupo Australian Research Data Commons (ARDC)¹⁵⁰ y la publicación de pautas sobre cómo aumentar el acceso a datos científicos para abordar los requisitos de un memorando de 2013 de la Oficina de Política de Ciencia y Tecnología elaboradas por las agencias de investigación de Estados Unidos (ARDC. Australian Research Data Commons, s.f.).
- Reconocer e incentivar las prácticas de los datos FAIR. Hasta ahora, el valor de una investigación científica se mide por el número de citas, aunque debería ser por el uso de los datos científicos, que en definitiva son los que tendrán un mayor impacto en la comunidad. Se plantea, también, la necesidad de que los investigadores reciban crédito y reconocimiento por el esfuerzo intelectual que supone proporcionar datos de valor, promoviendo así una “buena ciencia”.
- Financiar la infraestructura global para respaldar los datos y herramientas FAIR. Un tema central es la infraestructura tecnológica, donde la sostenibilidad de los repositorios de datos juega un papel de primera necesidad. Se establece que los investigadores no deben asumir el costo total del tránsito hacia los datos FAIR, al tiempo que se insta a una coordinación internacional de los gobiernos para que aquellos países con menos recursos sean beneficiados por los que tienen más en la conversión tecnológica de los repositorios de datos.

En el futuro, se espera que el proyecto OpenAIRE robustezca la colaboración entre lo social y lo técnico para consolidar la CA y el AA no sólo en Europa, sino en el resto del mundo. Para lograrlo, se trabaja en diversos protocolos de calidad y directrices respecto de los datos de investigación.

De allí que OpenAIRE entregue lineamientos claros respecto del AA (Spichtinger, s.f.):

- Obligación de proveer AA a través de las vías Verde o Dorada en todas las áreas (deposición obligatoria de cualquier manera).
- Embargos permitidos: 6/12 meses.

¹⁵⁰ Australian Research Data Commons (ARDC) es una iniciativa de transformación que permite a la comunidad de investigación australiana y a la industria el acceso a la infraestructura electrónica, las plataformas, las habilidades y las colecciones de datos de alta calidad de gran importancia nacional y de vanguardia. Sitio web: <https://ardc.edu.au/>

- Costos de AA Dorado elegibles para reembolso como parte del presupuesto del proyecto mientras el proyecto se ejecuta y apoyo posterior a la subvención pilotado a través de OpenAIRE.
- Se anima a los autores a conservar los derechos de autor y conceder licencias en su lugar.

Entre las más recientes iniciativas respaldadas por OpenAIRE se encuentra la EOSC, creada por la CE para promover el acceso y la reutilización de los datos de investigación que surgen de la investigación financiada con fondos públicos, contribuyendo a la implementación de políticas, flujos de trabajo e infraestructuras de CA en el contexto de un mercado único digital en Europa. “En la actualidad, existe un acceso fragmentado a los datos de investigación que existen, almacenan y se crean en muchos centros de datos, instituciones y centros de investigación diferentes en toda Europa. El AA a estos datos no es un hecho y el contenido no es interoperable, lo que restringe la investigación interdisciplinaria. EOSC resolverá este problema proporcionando un fácil acceso a estos datos, abriendo los datos financiados con fondos públicos, y además proveerá un único punto de acceso gratuito, asegurando que todas las bases de datos sean interoperables” (OpenAIRE, s.f.).

La idea es que esta iniciativa proporcione a la ciencia, a la industria y a las autoridades públicas europeas una infraestructura de datos de clase mundial para almacenarlos y gestionarlos, conectividad de alta velocidad para transportarlos (infraestructura europea de datos) y computadoras de alto rendimiento -cada vez más potentes- para procesarlos en la tecnología HPC¹⁵¹. Desde 2017 se han implementado planes de la puesta en marcha de la nube, al tiempo que se han elaborado orientaciones estratégicas y acciones concretas para la nueva estructura de gobernanza la EOSC (OpenAIRE, s.f.).

Concordante con lo anterior es la postura de Cathrin Stöver, CCO GÉANT y Vicepresidenta de la Junta Ejecutiva EOSC, en Alemania, quien en el contexto de la reciente Conferencia TICAL2020/ 4° Encuentro Latinoamericano de e-Ciencia¹⁵², manifestó en su ponencia la necesidad de impulsar una mejor ciencia en AA a través de la infraestructura de la nube, fomentando una mirada multidisciplinaria y la colaboración a nivel internacional para seguir avanzando. Para la alta ejecutiva, el cambio de mentalidad de los estamentos involucrados en la investigación científica (AA, CA, datos de investigación y repositorios de datos) debe darse conjuntamente con el cambio tecnológico, sobre todo desde las esferas académicas para irradiar a otros ámbitos sociales.

Cathrin Stöver abogó también por los esfuerzos de la CE para generar investigación científica multidisciplinaria con datos integrados, donde -incluso- se involucren los estudiantes en las instituciones de educación superior, ya sea usando esta información o aportando nuevo

¹⁵¹ High Performance Computing (HPC o Informática de Alto Rendimiento, en español) es utilizada por científicos, ingenieros e investigadores para tareas informáticas de complejidad mayor, como las que se usan para la predicción meteorológica, la exploración de petróleo y gas, la física, la mecánica cuántica y otras áreas. Sitio web: <https://www.intel.es/>

¹⁵² Conferencia TICAL2020/ 4° Encuentro Latinoamericano de e-Ciencia.03-09-2020. Sitio web: https://www.youtube.com/watch?v=o2MSVkbRdWo&list=PLsPfupT39gzO7l1xMfefa_scmlhvpCXTN&index=6

conocimiento. Se refirió también al desarrollo de servicios globales y al fomento de los Principios FAIR en investigaciones científicas, reforzando su importancia para la universalidad de los repositorios de datos.

Además, expuso sobre los tres objetivos de las prioridades estratégicas de la EOSC: enseñar y recompensar las prácticas de AA y habilidades; normar los estándares, herramientas y servicios con el fin de permitir a los investigadores encontrar, acceder, reutilizar y combinar resultados; y generar una infraestructura sustentable y federada que permita el intercambio abierto de resultados científicos. Todo lo anterior teniendo como base los principios como co-creación, investigación dirigida, impulso desde la comunidad, diseño flexible, extensible y escalable, incremental e interactivo, práctico y participativo, participación continua y consultas/pruebas de usuarios (Amaro, 2020).

Las iniciativas descritas se enmarcan en el contexto del Espacio Europeo de Investigación (EEI o ERA, en inglés) (EC. European Commission, s.f.), por medio del cual los países de la UE buscan fortalecer sus bases científicas y tecnológicas, su competitividad y su capacidad para abordar colectivamente los grandes desafíos del mañana, propiciando el flujo libre de las investigaciones científicas.

En dicho sentido, hay que destacar también la iniciativa del Mercado Único Digital (DSM, en inglés), una de las diez prioridades políticas de la CE y una de las siete iniciativas de la estrategia Europa 2020, debido a que forma parte de la Agencia Digital para Europa (DAE, en inglés). Para el DSM se trata de garantizar “la libre circulación de bienes, personas, servicios y capitales (...) en condiciones de competencia leal y un alto nivel de protección de los datos personales y de los consumidores, independientemente de su nacionalidad o lugar de residencia” (EC. European Commission, s.f.).

Otra de las iniciativas de la CE que se relaciona con OpenAIRE es el Programa de Investigación e Innovación Horizonte 2020, cuyos proyectos participan en el piloto de datos de investigación abiertos, debiendo depositar los siguientes datos en un repositorio (OpenAIRE, s.f.):

- Todos los datos necesarios para validar los resultados presentados en las publicaciones científicas, incluidos los metadatos que describen los datos de investigación depositados (o “datos subyacentes”), labor que debe realizarse a la mayor brevedad posible.
- Cualquier otro dato (datos curados que no sean directamente atribuibles a una publicación, o datos brutos), incluidos los metadatos asociados, según se especifique, y dentro de los plazos establecidos en el Plan de Gestión de Datos (según el criterio individual de cada proyecto / beneficiario).
- Los proyectos deben proporcionar información a través del repositorio elegido sobre las herramientas que se necesitan para validar los resultados, por ejemplo, software especializado o código de software, algoritmos y protocolos de análisis (lo ideal es que los mismos proyectos proporcionen estos instrumentos o, alternativamente, proporcionen acceso directo).

6.2. América Latina

Si bien en América Latina hay una realidad alejada de lo que hoy ocurre en Europa, lo cierto es que hay plena consciencia de que es imprescindible unirse a OpenAIRE, convocatoria en la que fue categórica Bianca Amaro, Presidenta de LA Referencia y Coordinadora del Programa Brasileño de CA, IBICT, Brasil, durante la Conferencia TICAL2020 / 4° Encuentro Latinoamericano de e-Ciencia¹⁵³.

Su análisis partió de la premisa de que la ciencia es un bien público que debe beneficiar a las personas, independientemente del aspecto comercial que ésta conlleva (al respecto, citó el tema de la actual pandemia COVID-19 y cómo la sociedad espera resultados beneficiosos a partir de las investigaciones científicas). También expresó que la comunicación es vital en el desarrollo científico y que una investigación no puede darse por concluida hasta su publicación y evaluación por parte de otros investigadores, quienes pondrán la mirada en los datos de investigación sobre los cuales aquélla se sustenta.

Bianca Amaro destacó la importancia de internacionalizar el conocimiento científico, tema del que, en su opinión, hay una mirada parcial de parte de algunas instituciones académicas latinoamericanas, pese al desarrollo de investigaciones científicas y a un sinfín de oportunidades para aportar a la comunidad científica. Con todo, desde sus inicios en 2012 con seis países, LA Referencia ha crecido hasta alcanzar hoy diez países integrantes: Argentina, Brasil, Colombia, Chile, Costa Rica, Ecuador, El Salvador, México, Perú y Uruguay. Juntos, suman alrededor de 2 millones 100 mil documentos en el repositorio LA Referencia, cosechados de los organismos encargados de realizar esta misma acción dentro de las fronteras de cada país (LA Referencia, s.f.). Según manifestó, “queremos ser un puente de conexión con el mundo en general y hemos avanzado en términos de red”, para lo cual continúan en conversaciones con más países con vistas a integrarlos (de hecho, anticipó que hay dos con los cuales se ha avanzado mucho al respecto) (Amaro, 2020).

¹⁵³ Conferencia TICAL2020 / 4° Encuentro Latinoamericano de e-Ciencia. 03-09-2020. Sitio web: https://www.youtube.com/watch?v=o2MSVkbRdWo&list=PLsPfupT39gzO7l1xMfefa_scmlhvpCXTN&index=6



Figura 6-1: LA Referencia en América Latina

Fuente: <http://www.lareferencia.info/es/>

Bianca Amaro puntualizó que la red se sustenta en RedCLARA¹⁵⁴ y que sus servicios están centrados en tres postulados: acuerdos, directrices y tecnología. De hecho, destacó su alineamiento con OpenAIRE, esforzándose por implementar todas las indicaciones emanadas desde aquí, por lo que trabajan con datos de investigación científicos curados y bajo la modalidad de cosecha. Esto quiere decir que las investigaciones de un determinado país son agrupadas en nodos nacionales, los cuales son cosechados o agregados al repositorio de datos de LA Referencia, cuya producción general es cosechada -a su vez- por OpenAIRE, con lo cual la investigación y los datos que ésta contiene alcanzan una mayor visualización, impactando positivamente en la CA y generando un efecto sobre la sociedad, impensado en el modelo anterior.

¹⁵⁴ RedCLARA es una organización internacional cuyo objetivo es conectar las redes informáticas académicas de América Latina. Sitio web: <https://www.redclara.net/index.php/es/somos>

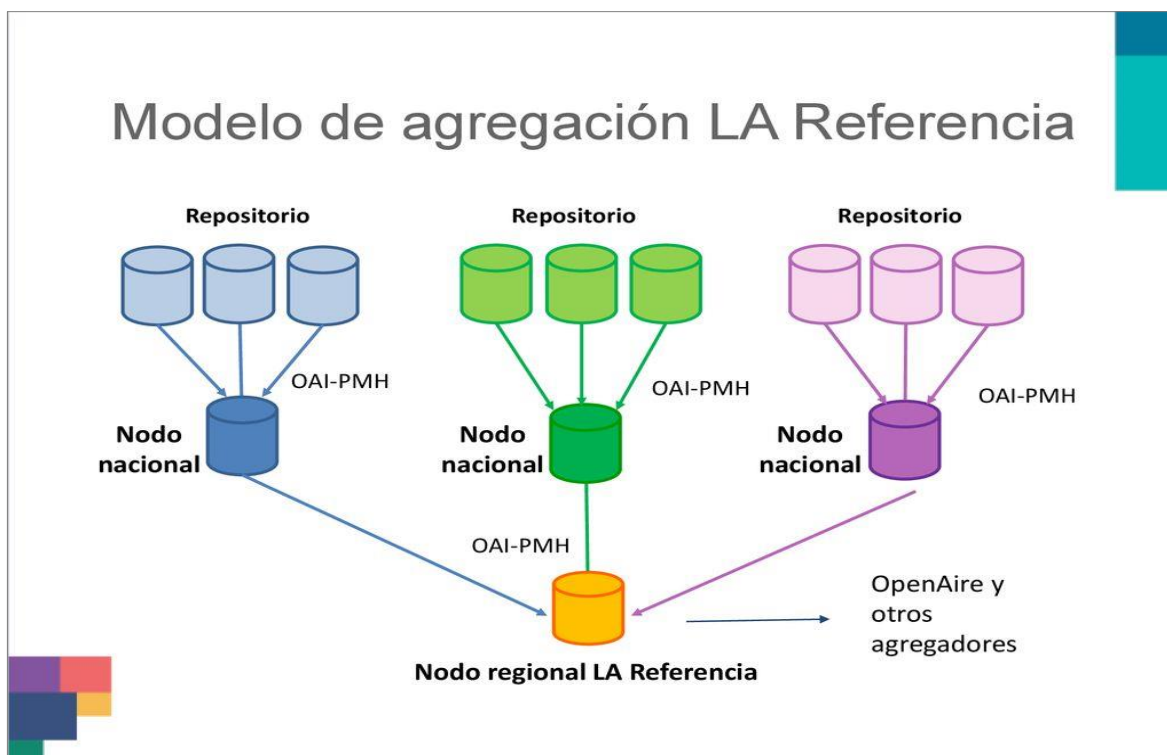


Figura 6-2: Modelo de agregación de LA Referencia

Fuente: <https://slideplayer.es/slide/17141730/>

Si bien constató los esfuerzos que LA Referencia viene realizando desde sus orígenes, expresó que aún hay una brecha importante con respecto a Europa, donde en materia de legislación sobre CA, AA y repositorios de datos se ha avanzado a un punto que parece demasiado lejano de alcanzar por parte de nuestro continente, donde apenas 3 países han realizado contribuciones en esa materia, como son Argentina, México y Perú. A su juicio, urge generar políticas de país que incentiven seriamente la producción de investigación científica porque, bajo las directrices de OpenAIRE, no sólo se acelera y optimiza la investigación, sino que también aumenta el rigor científico (resultados son validados por la veracidad de los datos científicos en que se sustenta), permite su reproducción, elimina la pérdida de datos, ahorra recursos económicos que pueden ser reorientados a otras investigaciones y hace que los datos de investigación estén disponibles para otros investigadores en el mundo, dando cuenta en definitiva de los Principios FAIR.

Según cifras de LA Referencia, en “Latinoamérica la mayoría de los datos no dialoga, los datos se pierden y/o son difíciles de encontrar, con lo que se limita la expansión del descubrimiento efectivo del conocimiento”. Se estima que el 20% de los enlaces de datos suplementarios se “degradan” cada año, el 80% de los datos se pierde, el 60% de los datos de la literatura científica permanecen inaccesibles y cerca del 50% de las investigaciones no son reproducibles.

Por eso, se ha planteado como desafío convencer a las entidades financiadoras a vincular esta acción al AA para las publicaciones y sus datos, convencer a los investigadores de que sus datos son

públicos si el financiamiento también lo es y convencer a los gestores de la importancia de apoyar la creación de una infraestructura para la CA, donde no sólo se consideren los aspectos tecnológicos (como software de repositorio, por ejemplo), sino también la capacitación del recurso humano que trabajará con estos temas. Otros desafíos que mencionó se relacionan con la introducción de un cambio en la forma de evaluar a los investigadores, más acorde con las directrices de CA, AA y datos de investigación según los criterios de OpenAIRE, en sintonía con los Principios FAIR; lo mismo que el tratamiento de cuestiones éticas y legales; la integración de los repositorios de publicaciones a los repositorios de datos; y la actuación en bloque (Amaro, 2020).

6.3. Brasil

Sin duda, uno de los países en el Cono Sur que más ha avanzado en términos de CA, AA y repositorios de datos de investigación es Brasil, según el análisis realizado por Washington Segundo, Coordinador del Laboratorio de Metodología de Tratamiento y Difusión de Información (COLAB), en el Instituto Brasileño de Información en Ciencia y Tecnología (IBICIT/ MCTIC), durante su participación en el Webinar de RedCLARA Cambio Climático “Repositorios Científicos de Datos Abiertos”¹⁵⁵.

Según el análisis del experto, buena parte de la producción científica generada en el país (como tesis de maestrías y doctorados, más capítulos de libros, libros, entre otros) se encuentra alojada en repositorios de publicaciones, con una evidente preponderancia de contribuciones por parte de las ciudades más grandes y modernas del país, ubicadas preferentemente en el sudoeste, como son Sao Paulo y Río de Janeiro. Del total de repositorios, 135 -equivalentes al 90%- utilizan el software DSpace¹⁵⁶.

¹⁵⁵ Webinar de RedCLARA Cambio Climático “Repositorios Científicos de Datos Abiertos”. 14-05-2020. Sitio web: <https://www.youtube.com/watch?v=NmjCP7NGdss>

¹⁵⁶ Presentación “Experiencias de Brasil en la creación de una red de repositorios de datos de investigación”. Segundo, Washington. IBICIT/ RedCLARA (11-05-2020).

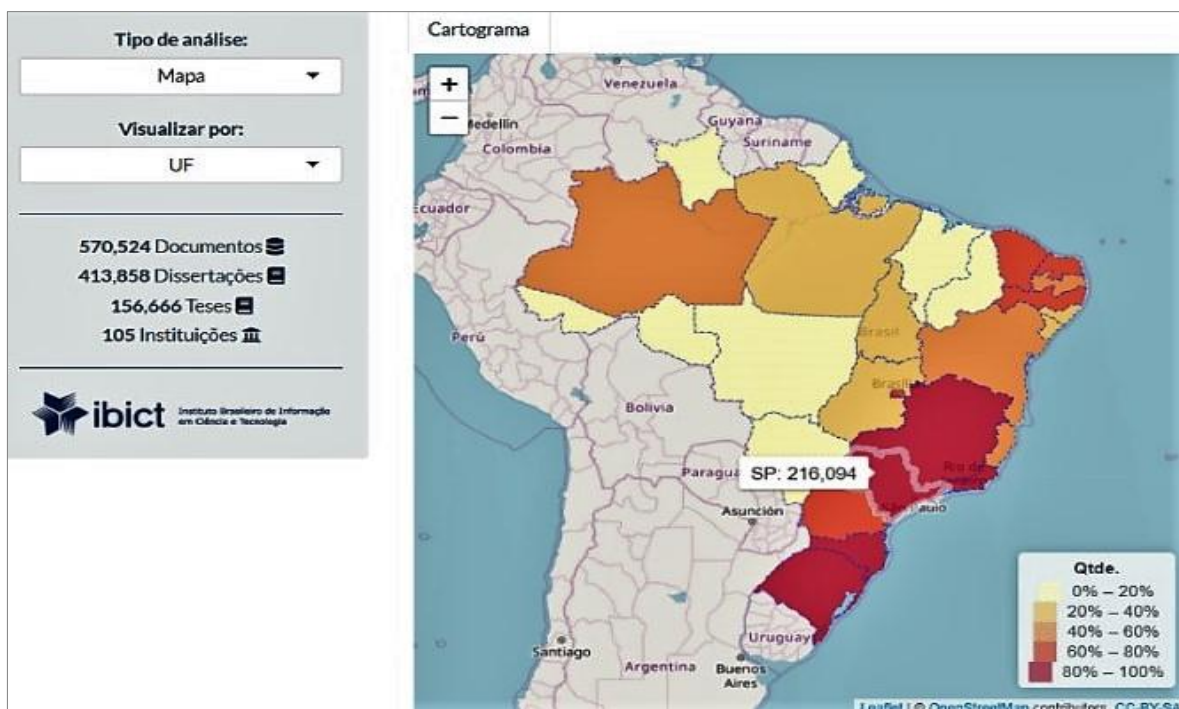


Figura 6-3: Distribución geográfica de publicaciones y datos en Brasil

Fuente: https://www.researchgate.net/figure/Cartogram_fig4_341129201

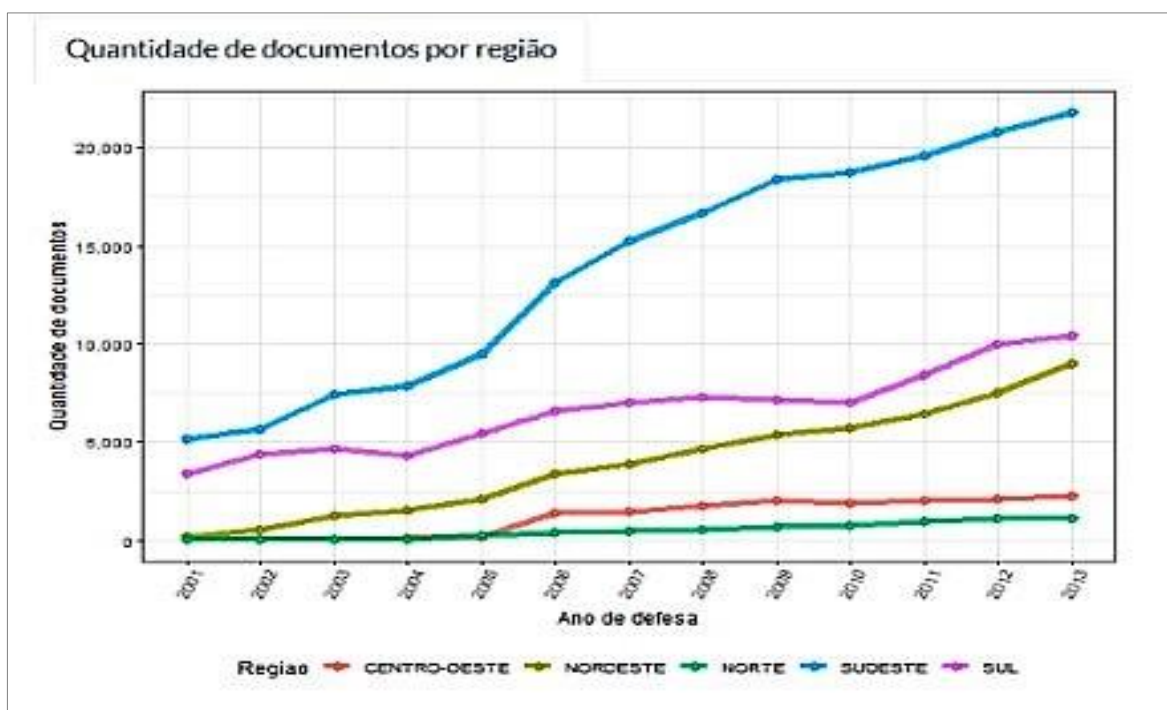


Figura 6-4: Producción de investigación por regiones en Brasil

https://www.researchgate.net/figure/Publications-per-year-by-geographic-group_fig3_341129201

Sin embargo, Washington Segundo es de la opinión de utilizar otro software para datos de investigación. Su juicio se basa en un estudio comparativo (Rufino Campêlo, 2020) de los softwares Dataverse, INVENIO, CKan y DSpace, exhibido en su ponencia, donde se mostró que éste último no estaba preparado para alojar adecuadamente los datos de investigación, porque requería de un desarrollo adicional. En cambio, Dataverse resultó más maduro y pro, garantizando atributos como dimensión, descripción, formato y ciclos de vida, además de contar con el respaldo de miles de usuarios en el mundo.



Pontos Observados	Programado Na linguagem	Ferramenta de pesquisa	Interface gráfica	Banco de Dados	Comunicação De interfaces	Processo de instalação	Formatos atendidos
Dataverse	JAVA	Solr	PrimeFaces E Bootstrap	PostgreSQL	OAI -PMH	Fácil, documentação Bem detalhada.	SPSS, STATA, R, XLSX, CSV
Invenio	Flask (Python)	Elasticsearch	WSGI, e Aipo	PostgreSQL	API REST	Complicada, Instalação incompleta	IIIF Image API support
DSpace	JAVA	Solr	JSPUI, XMLUI	PostgreSQL	OAI -PMH	Fácil, Documentação Bem detalhada.	PDF, XML, txt, asc, MARC, JPEG, JPG, GIF, PNG, TIFF, AIFF, RTF, Postscript
CKAN	Python	Solr	ckan plugins	PostgreSQL	Not enabled	Fácil, Documentação Bem detalhada..	XML and JSON CSV, Excel, XML, PDF, RDF

Figura 6-5: Razones por las cuales Brasil eligió a Dataverse

Fuente: Apresentação “Experiências de Brasil na criação de uma rede de repositórios de dados de investigação”. Segundo, W. IBICT/ RedCLARA. 11-05-2020.

Siguiendo con lo expuesto en su conferencia, el objetivo en el que ya se encuentran trabajando en Brasil consiste en la generación de una infraestructura que permita no sólo compartir y reutilizar los datos de investigación, sino también garantizar atributos como la curación, visibilidad, gestión, preservación y ahorro de recursos, lo que en definitiva se traduciría en una buena cosecha de datos, permitiendo ahorrarlos para nuevas investigaciones, cercanas en el tiempo y similares en el contenido. En dicho sentido, su mirada coincide con LA Referencia, de la cual Brasil es un nodo.

Para alcanzar esta realidad, el también Doctor en Informática sostuvo que hace años se debió implementar un modelo centralizado, donde -en un mismo Estado federativo- las instituciones mayores cosechaban los datos de investigación generados por otras menores, siendo aquéllas cosechados nuevamente por otras instituciones supraestatales, con el consiguiente extravío o

desaprovechamiento de los datos. De aquel modelo se evolucionó al actual, de tipo descentralizado, desde donde están migrando hacia el modelo distribuido¹⁵⁷.

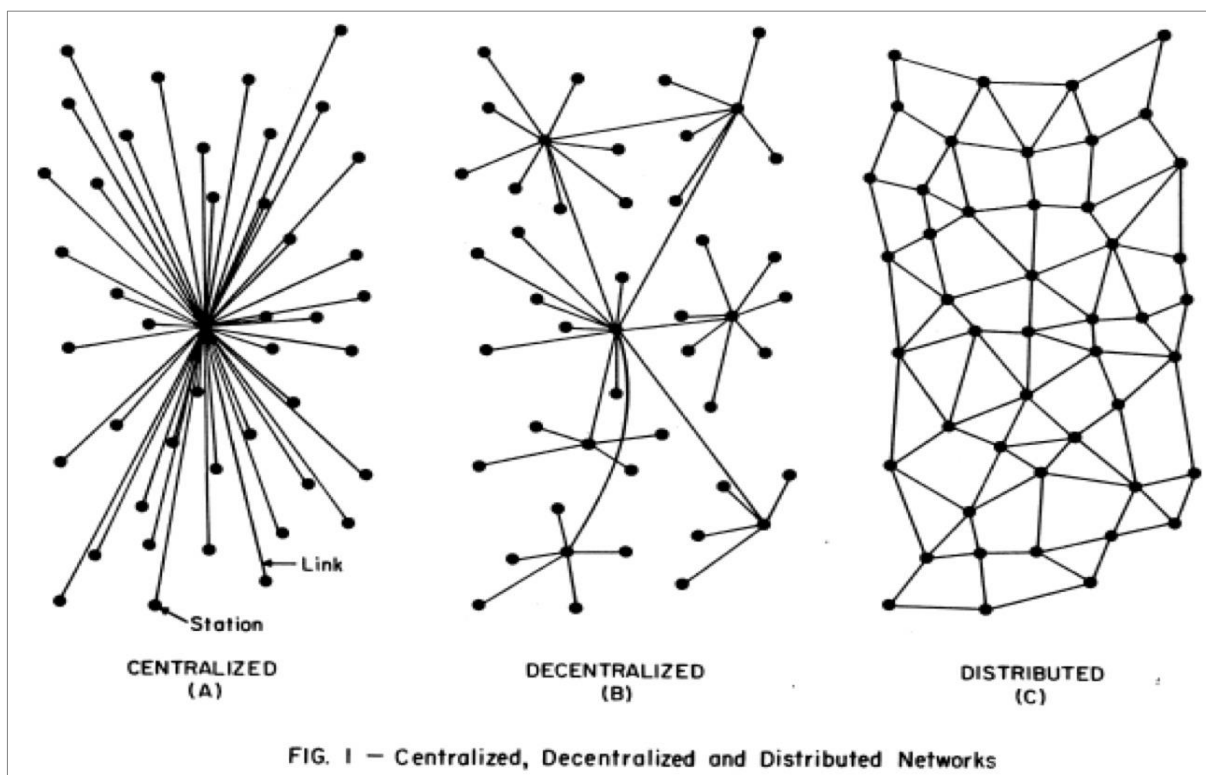


Figura 6-6: Modelos de redes de repositorios

Fuente: <https://medium.com/distributed-economy/what-is-the-difference-between-decentralized-and-distributed-systems-f4190a5c6462>

Básicamente, la diferencia entre ambos modelos radica en dos aspectos: en primer lugar, cómo y dónde se toma la decisión de subir los datos y, en segundo, cómo ésta se comparte entre los nodos del sistema. Mientras que en el actual modelo descentralizado cada institución gestiona su información, careciendo de una mirada completa del sistema, en el modelo distribuido se garantiza que el procesamiento de los datos se comparte por igual entre todos los nodos de la misma red, aunque las decisiones y parámetros se encuentran centralizados para efectos del sistema en general (Eagar, 2017). De esta forma, se garantizaría una cosecha ciento por ciento eficiente, con posibilidad de recuperar datos, mayor visibilidad e impacto real.

En su conferencia, el experto afirmó que con este modelo se estimularía a los investigadores para compartir sus datos de investigación en Oasisbr¹⁵⁸, priorizando este repositorio nacional común en vez de optar por publicar directamente en repositorios foráneos como Figshare, Apollo, GEO, Plos, Zenodo y NIH. Esta nueva política tiene la ventaja adicional de que, siendo nodo de LA Referencia,

¹⁵⁷ Presentación “Experiencias de Brasil en la creación de una red de repositorios de datos de investigación”. Segundo, Washington. IBICT/ RedCLARA (11-05-2020).

¹⁵⁸ Portal brasileño de publicaciones científicas en Acceso Abierto. Sitio web: <http://oasisbr.ibict.br/vufind/>

los datos de investigación cosechados por esta plataforma (además de RCAAP¹⁵⁹), son nuevamente cosechados por OpenAIRE, abriendo las puertas a la CA, AA y datos de investigación en todo el mundo.

En resumen, el investigador puntualiza que la apuesta de Brasil en materia de repositorios de datos es generar un federado (de corte general), a partir de 4 repositorios base: IBICT, CNPq, FIOCRUZ, Embrapa. Estos se vincularían entre sí a través de DataCite, teniendo a Dataverse como software oficial. En este mega repositorio se incluirían, también, instituciones que hoy no cuentan con sus propios repositorios. En dicho sentido, el IBICT desempeña el rol no sólo de soporte para la creación de repositorios de datos de investigación, sino también de desarrollo, adaptación y difusión de las directrices, al igual que de metabuscador de conjuntos de datos. Acorde con esta política, han implementado dos acciones de suma importancia, como la creación de una Guía de Ayuda en GitHub¹⁶⁰, y un grupo de trabajo/ayuda en Dataverse¹⁶¹.

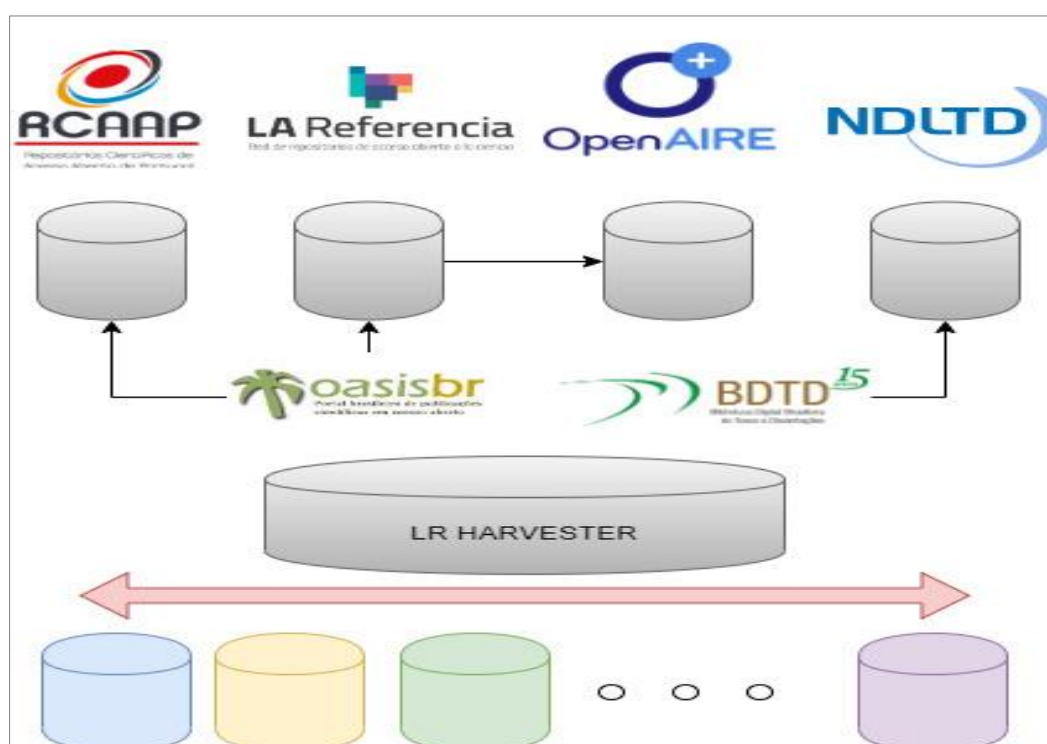


Figura 6-7: Modelo de cosecha de repositorios de datos de investigación en Brasil

Fuente: Presentación "Experiencias de Brasil en la creación de una red de repositorios de datos de investigación". Segundo, W. IBICT/ RedCLARA. 11-05-2020.

¹⁵⁹ Repositorios Científicos de Acceso Abierto de Portugal (RCAAP) es una iniciativa de Acceso Abierto creada en 2008 y promovida por la Agencia de la Sociedad del Conocimiento y operada por la Fundación para la Ciencia y la Tecnología, contando con el apoyo de la Universidad de Minho. Sitio web: <http://projecto.rcaap.pt/>

¹⁶⁰ GitHub plataforma de desarrollo colaborativo que se fundó en 2007, aunque al año siguiente comenzó sus operaciones. Se utiliza principalmente para la creación de código fuente de programas de ordenador y se considera una de las plataformas más importante de colaboración para proyectos de Acceso Abierto. Sitio web: <https://github.com/>

¹⁶¹ Presentación "Experiencias de Brasil en la creación de una red de repositorios de datos de investigación". Segundo, Washington. IBICT/ RedCLARA (11-05-2020).

6.4. Chile

Al igual que en el resto de Latinoamérica, en nuestro país el tema del desarrollo y expansión de los repositorios de datos de investigación es, también, incipiente, aunque se está consciente de la importancia que tiene -en términos de CA, AA y datos de investigación- compartir las investigaciones y datos en una comunidad globalizada, como es OpenAIRE, que cosecha publicaciones nacionales a través de LA Referencia, de la cual Chile es también un nodo.

“Hasta ahora, el resultado de las investigaciones científicas se publica, principalmente, a través de papers o bases de datos, a las que se accede mediante una modalidad de pago. Es por ello que, desde hace años, agencias dedicadas a financiar ciencia en todo el mundo se han planteado la forma de poner dicho conocimiento, generado a través del financiamiento estatal, a disposición de la comunidad científica en general, y de la ciudadanía”, explica Aisén Etcheverry, Directora Nacional de ANID (ANID. Agencia Nacional de Investigación y Desarrollo, 2020).

Con el fin de sondear la postura respecto a la nueva política de AA por parte de diversos organismos, actores y expertos en la materia, tanto en Chile como en el extranjero, la ANID realizó entre el 1 de junio y el 31 de julio de 2020 una consulta pública cuyos resultados serán cruciales para las decisiones que se implementarán a partir de 2021. En su desarrollo e implementación, esta herramienta contó con la participación de la Subdirección de Redes, Estrategia y Conocimiento (REC).

“Por primera vez ANID realiza una consulta de estas características, como una nueva forma de conocer las opiniones de todos los actores de la sociedad sobre este tema, especialmente, de parte de quienes los utilizan”, señala Patricia Muñoz, Directora de REC (ANID. Agencia Nacional de Investigación y Desarrollo, 2020).

Al momento del cierre de la encuesta (que midió 44 variables), el pasado 31 de julio, se contabilizaron 736 respuestas recibidas, como se desprende del documento oficial publicado en noviembre pasado. Según los primeros análisis del Departamento de Estudios de la ANID, estos son los principales datos (ANID. Agencia Nacional de Investigación y Desarrollo, 2020):

- Tipo de persona:
 - Jurídica: 2%
 - Natural: 98%
- Procedencia de respuestas:
 - Región Metropolitana: 42,8%
 - Región de Valparaíso: 10,4%
 - Región del Biobío: 10,4%
- Participación según género:
 - Hombres: 60,6%
 - Mujeres: 37,2%

- Representantes por disciplina:
Ciencias Naturales: 33%
Ciencias Sociales: 32%
Medicina y Ciencias de la Salud: 13%
Ingeniería y Tecnología: 11%
Humanidades: 7%
Ciencias Agrícolas: 4%

En línea con los estatutos de OpenAIRE y de LA Referencia, la Directora Nacional de la ANID destacó que “queremos partir de la lógica de que todo conocimiento generado con fondos públicos debe reportar beneficios para la sociedad y, en consecuencia, estar disponible para su uso” (ANID. Agencia Nacional de Investigación y Desarrollo, 2020).

Por su parte, la Directora de REC enfatizó que esta encuesta se suma a un recorrido que comenzó hace 10 años, a través de SciELO-Chile¹⁶² y LA Referencia, con el fin de mejorar y sumarse a las directrices mundiales que sobre estas materias se promueven en el mundo, siendo lideradas por OpenAIRE: “Hemos pavimentado este camino a través de diversos instrumentos durante estos años, lo que nos ha permitido llevar adelante este trabajo, generando una cultura institucional respecto a la necesidad de propiciar el acceso abierto a la información científica” (ANID. Agencia Nacional de Investigación y Desarrollo, 2020).

Sobre su adhesión a la Política de Acceso Abierto que “busca asegurar la disponibilidad del conocimiento científico contenido en las publicaciones resultantes de la ejecución de proyectos de investigación financiados con recursos públicos”, incluyendo asimismo los resultados de las tesis producidas en el contexto de los programas de becas y los datos de investigación, la mayoría se mostró Muy de acuerdo/De acuerdo, como puede apreciarse en la siguiente imagen:

Escala	N	%
Muy de acuerdo	445	60,8%
De acuerdo	218	29,8%
Ni acuerdo ni en desacuerdo	21	2,9%
En desacuerdo	31	4,2%
Muy en desacuerdo	8	1,1%
No responde	6	0,8%
No sabe	3	0,4%
Total general	732	100,0%

Fuente: ANID, 2020.

¹⁶² SciELO Chile es una biblioteca electrónica científica que incluye un grupo de revistas chilenas seleccionadas y desarrollado por la Comisión Nacional de Investigación Científica y Tecnológica (CONICYT). Sitio web: <https://scielo.conicyt.cl/>

Figura 6-8: “Respecto del objetivo de esta política ¿Qué tan de acuerdo o en desacuerdo está usted?”

Fuente:

https://www.uahurtado.cl/wp-images/uploads/2020/11/ResultadosConsultaPublica_PAA_2020.pdf

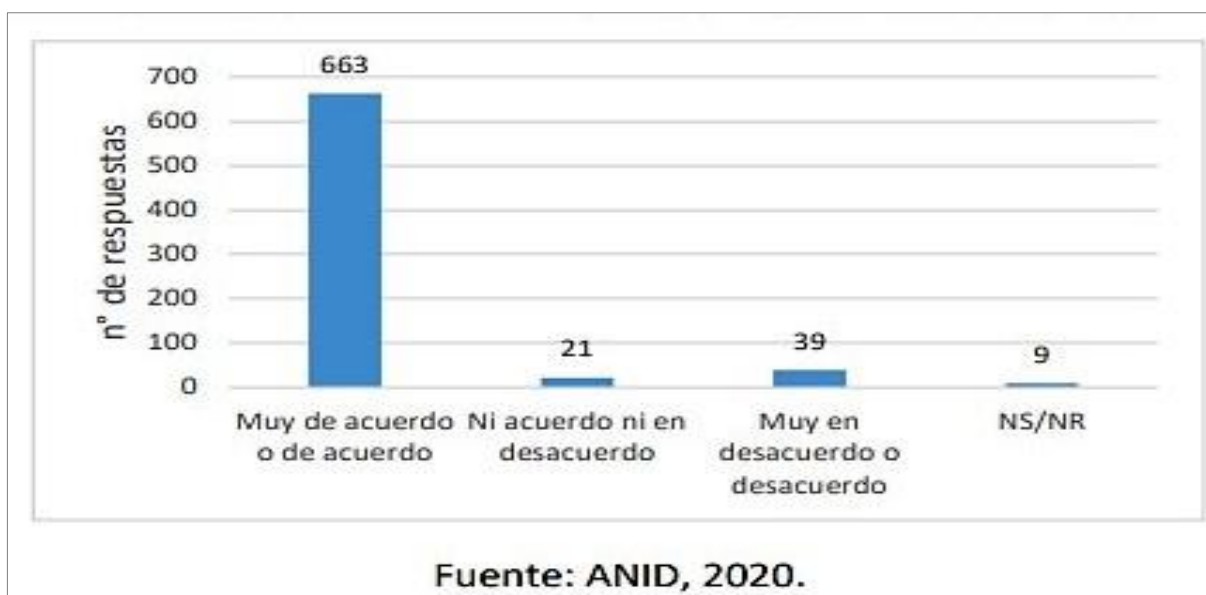


Figura 6-9: “Nivel de acuerdo/desacuerdo respecto al objetivo de la política”

Fuente:

https://www.uahurtado.cl/wp-images/uploads/2020/11/ResultadosConsultaPublica_PAA_2020.pdf

Respecto de las disposiciones de la Política de Datos de Investigación, la ANID plantea lo siguiente en términos generales (ANID. Agencia Nacional de Investigación y Desarrollo, 2020):

- Los investigadores deben presentar plan de gestión de datos de investigación con especificaciones propias de su área de conocimiento.
- Los investigadores deben depositar datos de investigación en repositorio de la ANID o de las instituciones de pertenencia.
- Estos repositorios deben cumplir estándares que garanticen el AA y resguardo de documentos y/o datos.
- Si los investigadores optan por un repositorio diferente a ANID, deben informar a ésta la URL de sus publicaciones o identificadores de éstas al igual que de los conjuntos de datos, teniendo como plazo máximo 1 año para hacerlo.
- Se excluyen del AA todos los datos de investigación que deban mantenerse en confidencialidad, informando previamente a ANID de dicho carácter.

Sobre este punto, la mayoría se declara estar Muy de acuerdo/De acuerdo, con el 71,4% de las preferencias. En cambio, un 14,7% se declara Muy en desacuerdo/En desacuerdo, ratificando la tendencia adoptada por los investigadores sobre compartir los datos de investigación en repositorios de AA, reflejada también en las políticas de LA Referencia y de OpenAIRE.

Escala	N	%
Muy de acuerdo	263	35,9%
De acuerdo	260	35,5%
Ni acuerdo ni en desacuerdo	71	9,7%
En desacuerdo	66	9,0%
Muy en desacuerdo	42	5,7%
No responde	13	1,8%
No sabe	17	2,4%
Total general	732	100,0%

Fuente: ANID, 2020.

Figura 6-10: “Indique su nivel de acuerdo o desacuerdo respecto a las disposiciones sobre datos de investigación que señala la Política de Acceso Abierto”

Fuente:

https://www.uahurtado.cl/wp-images/uploads/2020/11/ResultadosConsultaPublica_PAA_2020.pdf

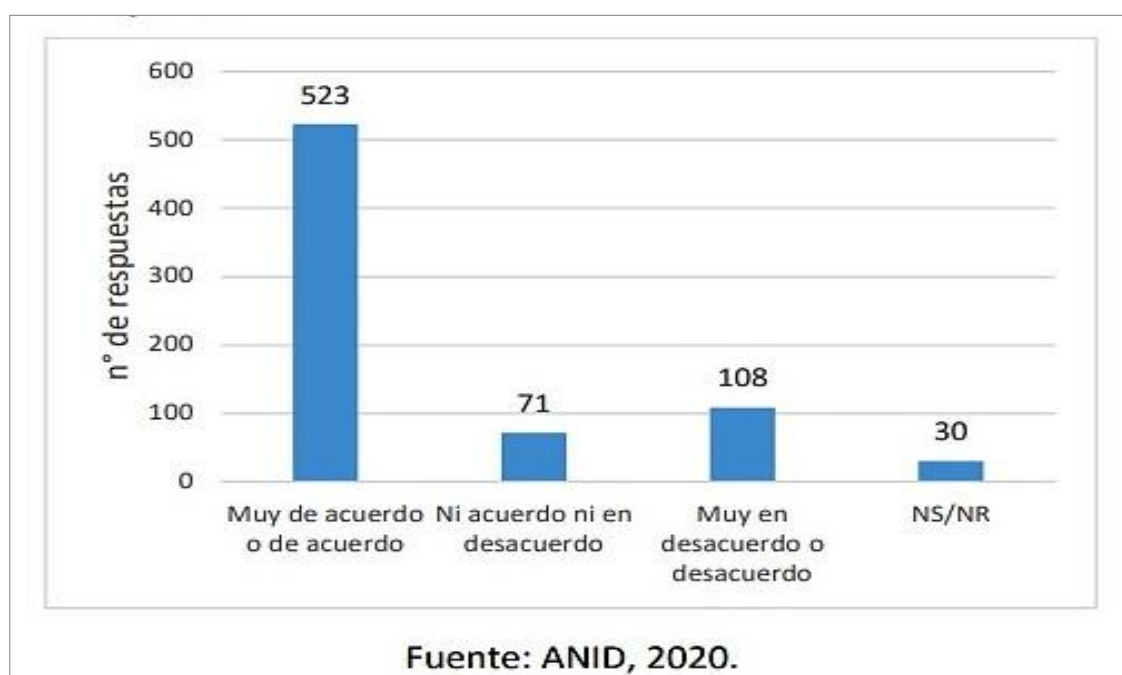


Figura 6-11: “Nivel de acuerdo/desacuerdo respecto a las disposiciones sobre datos de investigación de la política”

Fuente:

https://www.uahurtado.cl/wp-images/uploads/2020/11/ResultadosConsultaPublica_PAA_2020.pdf

Sobre la base de este mapeo, a lo que suman las iniciativas desarrolladas con anterioridad, la ANID trazó una hoja de ruta cuyo primer hito se materializará en 2021, a partir de lo realizado durante el último trimestre de 2020, donde se trabajó en un proyecto de desarrollo del repositorio para la Política de Acceso Abierto, el cual utilizará la más reciente versión del software DSpace, garantizando la interoperabilidad con otros servicios de información de la ANID como Dataciencia y el Portal del Investigador (ANID. Agencia Nacional de Investigación y Desarrollo, 2020).

Asimismo, desde abril de 2021, se trabajará en las capacitaciones de los investigadores y de todo el personal que se requiera para la implementación de esta política de AA, lo cual se hará en conjunto con las vicerrectorías de Investigación y Desarrollo, además de las direcciones de bibliotecas de las universidades. Se contará, también, con el apoyo del Ministerio de Ciencia, Tecnología, Conocimiento e Innovación. La modalidad de formación, capacitación y difusión será a través de instancias formativas virtuales y presenciales, considerando también materiales de apoyo (como tutoriales, vídeos y sesiones grabadas de webinars), disponibles en el sitio web de la ANID (ANID. Agencia Nacional de Investigación y Desarrollo, 2020).

Respecto de los aspectos legales, la ANID precisó que el próximo año también incluirán en la agenda estas materias con otros actores del sistema para “abordar todas las situaciones donde pueden existir colisiones de derechos o problemas de interpretación, incluidos los temas de uso de licencias abiertas, tanto para las publicaciones como para los datos”. Dicha iniciativa debería desembocar en la promulgación de un reglamento específico, entre 2021 y 2022, sobre datos de investigación, por medio del cual se asegure el cumplimiento de los Principios FAIR (ANID. Agencia Nacional de Investigación y Desarrollo, 2020).

Otro tema relevante para la ANID en 2021 se relaciona con las editoriales, consideradas entidades de alta relevancia en el circuito de la comunicación científica. En dicho sentido, ya se están solicitando acuerdos para trazar una Ruta Verde con las que participan en el Programa BEIC¹⁶³. “El objetivo es poder entregar a los autores instrucciones específicas para los temas del depósito, la cesión parcial de derechos, el uso de licencias Creative Commons y los plazos de embargo, y que ya estén consensuados con las editoriales” (ANID. Agencia Nacional de Investigación y Desarrollo, 2020).

En sintonía con lo anterior, y con el fin de facilitar los procedimientos ya descritos -al igual que las buenas prácticas-, la ANID elaborará manuales de apoyo para los planes de gestión con el fin de hacer más amigable el tránsito hacia esta nueva política de AA. Es así como para los datos de investigación existirá un repositorio, incentivando a los investigadores a depositarlos en repositorios temáticos o disciplinarios (ANID. Agencia Nacional de Investigación y Desarrollo, 2020).

¹⁶³ Biblioteca Electrónica de Información Científica (BEID) es el instrumento más importante de acceso a la información científica del país y es financiado por la ANID (ex CONICYT) bajo la administración de la Corporación CINCEL. Entrega acceso gratuito a través de Internet a un conjunto de revistas en texto completo. Sitio web: <https://www.beic.cl/>

7. Metodología

La metodología empleada en este trabajo se sustentó en seis instancias fundamentales:

- Revisión bibliográfica
- Encuesta
- Entrevista
- Comunicación y colaboración digital
- Análisis de dos casos de estudio
- Benchmarking de sitios web

Estas instancias, de carácter teórico y práctico, me permitieron realizar una búsqueda, selección y evaluación conducentes a lograr un acercamiento al tema y obtener material adecuado para elaborar y presentar la tesis que lleva por título “Implementación y gestión de un Repositorio de Datos de Investigación: Buenas Prácticas y Desafíos”.

A continuación, se presenta el alcance de cada una de las acciones utilizadas:

7.1. Revisión bibliográfica

La revisión bibliográfica del tema correspondió a una etapa teórica, la cual se elaboró íntegramente en fuentes de información en Acceso Abierto (AA), en un período comprendido entre 2017 y diciembre de 2020, con una cobertura idiomática en español y en inglés y con una cobertura geográfica que abarcó el estado del arte en esta materia en Europa, Estados Unidos, Canadá, Australia, Latinoamérica y Chile. Esta instancia requirió de la elaboración de estrategias de búsquedas, las que incluyeron conceptos y palabras clave, asociadas al tema en cuestión, a partir de los cuales tuvieron lugar nuevas búsquedas relacionadas. También se generaron “alertas” vinculadas al tema de interés en los recursos bibliográficos digitales revisados, con el propósito de acceder a lo más reciente.

Conceptos / Palabras clave:

Repositorios de datos de investigación, Repositorios de Próxima Generación, Datos de investigación, Ciencia Abierta, Acceso Abierto, Ciencia Ciudadana, Políticas de Acceso Abierto, Políticas de repositorios de datos, Licencias, Softwares para repositorios de datos, Desarrollo tecnológico de repositorios de datos, Buenas prácticas para repositorios de datos, Ciclo de vida de los datos de investigación, Certificaciones para repositorios de datos, Estándares, Identificadores persistentes, Metadatos, Directrices para repositorios de datos, Directorios de repositorios de datos de investigación.

La búsqueda de literatura se realizó en variados recursos disponibles a través del buscador en Bibliotecas UC¹⁶⁴ y, también, en Google Académico. En ambos casos se aplicaron filtros cronológicos (años de publicación 2017 a la fecha), por tipo de documentos y por coberturas geográficas e idiomáticas, con el objetivo de lograr especificidad en la recuperación de información y posterior revisión. Como mencioné anteriormente, y a modo de desafío, también se consideró que toda la bibliografía obtenida estuviese disponible en AA, con el fin de indagar -y posteriormente probar- que la literatura generada en AA y los recursos encontrados alusivos al tema son de calidad, proviniendo de fuentes de información revisadas por pares u originadas en instituciones conocidas. Como resultado, se obtuvieron preferentemente artículos, algunos libros, capítulos de libros y algunas tesis; no obstante, éstas últimas abordaban otro tipo de repositorios.

Por otra parte, la revisión bibliográfica del tema se extendió a directorios e instituciones académicas, de las cuales pude revisar algunas guías temáticas especializadas en la materia. Otras fuentes revisadas correspondieron a variados formatos, tales como seminarios y conferencias (webinars), entre las cuales me gustaría destacar las siguientes:

- “Conferencia TICAL2020 / 4º Encuentro Latinoamericano de e-Ciencia (03-09-2020)”: contó con ponencias de Bianca Amaro (Presidenta de LA Referencia y Coordinadora del Programa Brasileño de Ciencia Abierta, IBICT, Brasil), Cathrin Stöver (CCO GÉANT, Vicepresidenta Junta Ejecutiva EOSC, Alemania), María Esther Vidal (Jefa Grupo Investigación Gestión de Datos Científicos TIB, Venezuela), Kathleen Shearer (CEO COAR, Canadá) y Marcelo Volpi (Director Regional Cono Sur-CALA Enterprise-CommScope). Asimismo, participaron Luis Eliécer Cadenas (Director Ejecutivo de RedCLARA¹⁶⁵) y Luis Núñez (Físico, Universidad Industrial de Santander, Colombia).
- Webinar Cambio Climático “Repositorios Científicos de Datos Abiertos” (11-05-2020): con la conferencia “Experiencias de Brasil en la creación de una red de repositorios de datos de investigación”, de Washington Segundo, Coordinador del Laboratorio de Metodologías de Tratamiento y Difusión de Información (COLAB), del Instituto Brasileño de Información en Ciencia y Tecnología (IBICT / MCTIC); y con la conferencia “Cómo diseñar un proyecto institucional para la gestión de datos científicos abiertos”, de Rosalina Vázquez Tapia, Directora de la Biblioteca Virtual Universitaria (México), Coordinadora General de REMERI (Red Mexicana de Repositorios Institucionales) y Responsable Técnica del nodo nacional en LA Referencia. En tanto, la presentación estuvo a cargo de Laura Castellana, Coordinadora de Proyectos Académicos RedCLARA.

A la luz de lo anterior, se entrega una visión de las tendencias de desarrollo contemporáneas en esta materia, experimentado a nivel de continentes, de disciplinas y de aspectos de jurisprudencia, mediante las cuales es posible conocer y apreciar no sólo las similitudes, sino también las

¹⁶⁴ Sitio web: <https://bibliotecas.uc.cl/>

¹⁶⁵ RedCLARA es una organización internacional con sede en Montevideo (Uruguay), cuyo objetivo es conectar las redes informáticas académicas de América Latina para fortalecer el desarrollo de la ciencia, la educación, la cultura y la innovación a través del uso innovador de redes avanzadas. Sitio web: <https://www.redclara.net/>

diferencias, entre los distintos países en cuanto a la generación de políticas de AA y/o de datos de investigación. Finalmente, toda la bibliografía recuperada se encuentra también en esta publicación y fue citada con la Norma APA¹⁶⁶.

7.2. Encuesta

El objetivo de este instrumento fue conocer la realidad de las instituciones que acreditaban tener repositorios de datos de investigación con el propósito de recoger las principales prácticas, recomendaciones y experiencias de cada una de ellas para ponerlas a disposición de otras instituciones interesadas en mejorar su experiencia con los repositorios de datos de investigación, o bien, de instituciones que estén considerando implementar uno.

Por dicho motivo, este instrumento consideró la elaboración, aplicación y posterior análisis de los resultados obtenidos. Se desarrolló en español e inglés, considerando 15 preguntas en total, de las cuales 9 fueron cerradas (propuesta de respuesta entregada por el encuestador), mientras que 6 fueron abiertas (se pedía al encuestado entregar más antecedentes sobre algún tema en específico). De éstas últimas, las más significativas fueron N°14 y N°15, ya que ambas estaban orientadas no sólo a conocer la experiencia de las instituciones respecto de la implementación y desarrollo del software que eligieron para el repositorio de datos, sino también a obtener recomendaciones de buenas prácticas en la gestión del mismo. El criterio para la selección de las instituciones para encuestar se realizó a través de una búsqueda en el repositorio de datos de investigación re3data.org¹⁶⁷, privilegiando la elección de repositorios temáticos, en AA y con alguna certificación (deseable). Se buscaron instituciones de perfil académico y de investigación. La plataforma utilizada para desarrollar e implementar la encuesta fue LibWizard¹⁶⁸, de Springshare¹⁶⁹.

Con el objetivo de enriquecer este instrumento, y acogiendo una sugerencia de la profesora guía, en la elaboración de la encuesta y posterior análisis de resultados, participaron un ingeniero experto en proyectos tecnológicos y un periodista especializado en comunicaciones institucionales y contenidos web. La encuesta tuvo dos envíos, ya que la tasa de respuesta esperada era de 50, lo cual no se logró en la primera oportunidad, razón por la cual se preparó un segundo envío. Éste

¹⁶⁶ Sitio web: <https://guiastematicas.bibliotecas.uc.cl/normasapa>

¹⁶⁷ re3data.org es un registro global de repositorios de datos de investigación que cubre repositorios de datos de investigación de diferentes disciplinas académicas. Incluye repositorios que permiten el almacenamiento permanente y el acceso a conjuntos de datos a investigadores, organismos de financiación, editores e instituciones académicas. Sitio web: <https://www.re3data.org/>

¹⁶⁸ Es una plataforma del sitio Springshare que permite el uso de formularios, encuestas, cuestionarios y tutoriales con el fin de usarlos como herramientas de evaluación y comentarios para bibliotecas. Sitio web: <https://www.springshare.com/libwizard/>

¹⁶⁹ Empresa de tecnología orientada a la educación y fundada en 2007, cuyo objetivo es ayudar a las bibliotecas a brindar un servicio de calidad a los usuarios, proveyéndoles de softwares en la nube. Hoy, se encuentran presentes en 6 mil 100 bibliotecas en 82 países, con un alcance de 130 mil bibliotecarios y 20M+ de usuarios de biblioteca. Sitio web: <https://www.springshare.com/>

último se dirigió a instituciones que contaban con repositorio institucional, aunque -hay que precisar- no necesariamente contuviera una colección para datos de investigación.

Tabla 7-1: Resumen encuesta

Período de envío de la encuesta	13 al 15 de octubre de 2020 (primer envío) 29 de octubre de 2020 (recordatorio)	3 de noviembre de 2020 (segundo envío)
Destinos	117 instituciones	218 instituciones
Zonas geográficas	17 América Latina (2 Chile) 25 Australia 13 España 25 Estados Unidos y Canadá 25 Reino Unido 12 Unión Europea	185 América Latina (incluido Chile) 20 Estados Unidos 12 Unión Europea
Total de respuestas	52	
Tasa de respuestas	16%	

Para efectos de un análisis geográfico general, se utilizó la plataforma Who.is¹⁷⁰, la cual permitió identificar las siguientes regiones geográficas, con representación de instituciones que adhirieron a la invitación, pero cuyas respuestas son reservadas:

- Estados Unidos
- Canadá
- Reino Unido
- España
- Unión Europea
- Australia
- América Latina
- Chile

En relación al análisis y presentación de los resultados, se emplearon las plataformas LibWizard, de Springshare –como lo había señalado-, y también Excel, de Microsoft. De ambas aplicaciones se extrajeron gráficos. En el primer caso, los llamados gráficos de “mesa”, mientras que, en el segundo, los gráficos de barras. La razón por la cual se utilizaron ambos gráficos ha sido, exclusivamente, para guiar de mejor modo el detalle de los resultados.

Al contrastar ambos tipos de gráficos surgen algunas aparentes discrepancias numéricas debido a las siguientes razones: LibWizard sólo contabiliza las respuestas válidamente emitidas, asignando así puntuación (o vinculando) la respuesta a la pregunta sólo cuando aquella se ajusta a las

¹⁷⁰ Plataforma que facilita herramientas de búsqueda, nombre de dominio y sitio web, más localización geográfica de una IP. Sitio web: <https://who.is/>

propuestas como posibles opciones. En caso de que una respuesta no se ajuste el modelo preestablecido, la asigna en la categoría “Otros”, validando igualmente la respuesta. También valida la respuesta cuando es de desarrollo (solicitado por el encuestador). Sin embargo, si no hay respuesta por parte del encuestado, esto no se valoriza para efectos de conteo, con lo cual no es considerada como una respuesta válida, como se verá al comentar en detalle los resultados de la encuesta, en el próximo capítulo. En cambio, Excel valida numéricamente todas las respuestas, incluso aquellas que quedaron en blanco, a las que para efectos del presente estudio se les denominó “Sin Respuesta (s/r)”. Al igual que en la plataforma LibWizard, las respuestas entregadas que no se ajustaron a las definidas por el encuestador se agruparon en la categoría “Otros”. En los casos en que se entregaron más respuestas de las solicitadas por el encuestador, en la modalidad de opciones, se decidió validarlas igualmente, debido a que no distorsionaron la medición estadística de la encuesta, debido al bajo número de casos en que ocurrió lo antes descrito.

De esta forma, hay 9 casos en que se presentan ambos gráficos (de mesa y de barra), en 4 de los cuales las 52 respuestas fueron validadas por ambas plataformas, mientras que hay 5 casos en que las 52 respuestas fueron validadas con los criterios ya descritos, tanto por parte de LibWizard como por parte de Excel.

Por último, hay que señalar que la encuesta se sustentó sobre la base del criterio de absoluta confidencialidad para los participantes. Sólo se consideraron las respuestas para la elaboración de los gráficos, con el propósito de describir tendencias, conductas o realidades, al igual que de compartir -a través de la presente publicación- buenas prácticas con otras instituciones que se encuentren en la etapa de implementación o de mejoras de su repositorio de datos de investigación (lo que ocurre fundamentalmente con las dos respuestas abiertas, situadas al final de la encuesta). En el anexo 3 se encontrará el detalle de las preguntas en versión español e inglés.

7.3. Entrevista

Como tercer recurso, la metodología consideró una entrevista en profundidad, la cual fue realizada al destacado investigador mexicano Agustín Fernández Eguiarte. Este recurso no estuvo considerado en la génesis del proyecto, pero durante el desarrollo de éste, además de haber recibido sugerencias de incluirlo, me percaté de que era imprescindible contar con esta conversación. Dado que tuve la oportunidad de concretar un encuentro con un investigador, con conocimiento y experiencia sobre la tecnología de repositorios de datos, le solicité reunirnos para conversar al respecto. Como resultado, me siento satisfecha de haber compartido y recibido una información valiosa en cuanto a su visión y recomendaciones, así como una aproximación al futuro que él percibe en este ámbito.

¿Cómo contacté a este investigador? Su institución fue incluida en el contexto del primer envío de la encuesta, realizado en octubre pasado, luego del cual -sin esperarlo- estableció contacto conmigo vía correo electrónico.

Su postura, la cual logré percibir a través de los correos que mantuvimos para comunicarnos, me pareció interesante y significativa de rescatar para mi tesis, porque su experiencia corresponde a un proyecto relativamente nuevo, del cual conocemos el recorrido que siguió hasta convertirse en lo que es en la actualidad, dando cuenta también de aspectos tecnológicos y de recursos humanos necesarios para el desarrollo y buen funcionamiento de los repositorios de datos. Aspectos como la selección de un software según los objetivos del repositorio, la naturaleza de los datos de investigación que se alojarán en éste, la calidad de los datos propiamente tal, los criterios de almacenamiento, el AA y la adecuación con la normativa internacional sobre el desarrollo y gestión de esta plataforma eran contenidos a los cuales debía acceder. Del caso propiamente tal, se puede obtener el detalle completo en el capítulo 8.

Este encuentro/entrevista digital se realizó el 23 de noviembre, a través de la plataforma Zoom. Tuvo una extensión de 01 horas y 23 minutos. El diálogo giró en torno a 15 preguntas enviadas con anterioridad y relacionadas con las preguntas de la encuesta, aunque hubo leves ajustes sobre la base de cómo se fue dando el encuentro. La reunión virtual se programó a las 20:00 horas de Santiago y 17:00 horas de Ciudad de México. La grabación es mantenida bajo mi administración como testimonio cercano y de primera fuente; no obstante, la versión completa fue registrada como uno de los dos casos que conforman el capítulo 8 (ya mencionado), donde analizo pormenorizadamente dos experiencias relevantes respecto de la implementación y desarrollo de repositorios de datos de investigación, correspondiendo ésta a una en Latinoamérica.

7.4. Comunicación y colaboración digital

En relación a este punto, y como una forma de enriquecer los contenidos desarrollados en el capítulo “Situación Actual”, establecí contacto y solicité al investigador y Doctor en Informática, Washington Segundo, Coordinador del Laboratorio de Metodología de Tratamiento y Difusión de Información (COLAB), en el Instituto Brasileño de Información en Ciencia y Tecnología (IBICIT/MCTIC), apoyo bibliográfico referido básicamente a las infografías que exhibió en su intervención en el webinar Cambio Climático “Repositorios Científicos de Datos Abiertos”¹⁷¹, de RedCLARA.

Es así como, de vuelta de correo electrónico, no sólo me compartió la presentación completa ya aludida, en formato PDF, sino también otro archivo, en el mismo formato, correspondiente al estudio de dos investigadores brasileños sobre las características de los softwares más usados en la implementación y desarrollo de los repositorios, el cual tomó a su vez como referencia para su propia presentación en el mencionado webinar. Cabe señalar que este mismo estudio se encuentra disponible en la versión digital de una revista especializada de IBICT, enlace que también me compartió.

¹⁷¹ Webinar de RedCLARA Cambio Climático “Repositorios Científicos de Datos Abiertos”. 14-05-2020. Sitio web: <https://www.youtube.com/watch?v=NmjCP7NGdss>

Asimismo, el investigador me facilitó el enlace URL de otro artículo publicado por la psicóloga Mari Eagar en el sitio Medium¹⁷², en el cual se abordan las diferencias entre los sistemas descentralizados y distribuidos, y que él también utilizó como referencia bibliográfica para su propia exposición.

Cabe hacer notar que Washington Segundo me autorizó para utilizar estas imágenes con el fin de ser incluidas en mi tesis, del mismo modo que la bibliografía comentada anteriormente me fue referida para fundamentar de mejor manera los contenidos citados.

Utilizando este mismo método, y con resultados que me significaron la colaboración para articular uno de los casos presentados en mi tesis (capítulo 8), también tomé contacto con Andrew McKenna-Foster, Especialista de Producto de Figshare. Tanto de Washington Segundo como de Andrew McKenna-Foster quiero destacar que sus respuestas fueron ágiles y con aportes valiosos para los contenidos que abordé, lo que da cuenta de la relevancia que -para ambos expertos- tiene el compartir la información. A mi juicio, la actitud de ambos hacia mi tesis evidencia la coherencia entre la promoción teórica del AA y su materialización en la práctica.

7.5. Análisis de los casos de estudios

Como parte de la metodología, también consideré relevante presentar dos casos correspondientes a realidades diferentes, las cuales se apreciarán desde el estado de avance regional hasta las tecnologías empleadas para la implementación de repositorios de datos de investigación. Junto con describir las diferencias en la concepción de los mencionados casos, también se referirán concordancias en relación a recomendaciones y buenas prácticas, al igual que al interés y el ahínco de los mentores en llevar sus experiencias con un alto grado en cuanto al compromiso y a la calidad.

El primer caso de estudio corresponde a un “Repositorio de Datos Abiertos Científicos Geoespaciales” (Fernández Eguiarte, 2019), concebido como un Proyecto en Ciudad de México CdMx, específicamente en la Universidad Nacional Autónoma de México (UNAM). Este país -que a la fecha lidera en materia de repositorios de datos en nuestro continente- tiene a su haber una interesante experiencia materializada en el Repositorio Institucional del Centro de Ciencias de la Atmósfera, de la UNAM. Se trata de un Proyecto que nació en 2015 a cargo del destacado investigador Agustín Fernández Eguiarte (UNAM. Universidad Nacional Autónoma de México, 2017).

En tanto, el segundo caso aborda una experiencia exitosa de implementación de repositorios para datos de investigación en la Universidad Estatal de Iowa, utilizando la plataforma Figshare (Figshare, 2020), una de las más utilizadas en diversas instituciones de Europa y de Estados Unidos. Fundada en 2012 por Mark Hahnel, quien creó esta tecnología para repositorios mientras culminaba su Doctorado en Biología de Células Madre en el Colegio Imperial de Londres, se presenta como una

¹⁷² Medium es una plataforma abierta que tiene alrededor de 170 millones de lectores en el mundo, abierta a los profesionales de todas las disciplinas, quienes pueden publicar libremente artículos destinados a incentivar “un pensamiento dinámico y perspicaz”. Sitio web: <https://medium.com/>

infraestructura digital que tiene como inspiración la CA y el rol que deberían adoptar los investigadores. Hahnel ha concretado su predilección por este ámbito al ser integrante de DataCite¹⁷³ y del Directorio de Revistas de Acceso Abierto (DOAJ, en inglés)¹⁷⁴, además de asesor de Springer Nature (Figshare, s.f.).

Tanto el caso correspondiente al “Repositorio de Datos Abiertos Científicos Geoespaciales”, de la UNAM, como el caso asociado a una exitosa aplicación práctica del software Figshare, se desarrollarán en detalle en el próximo capítulo, **Aplicación a casos de estudio**.

7.6. Benchmarking de sitios web

Otra herramienta que utilicé en la metodología se relacionó con el benchmarking de sitios web, lo que me fue posible luego de consultar directorios mencionados también en esta publicación, a partir de los cuales efectué un levantamiento con los principales repositorios de datos de investigación internacionales que, actualmente, cumplen con la normativa vigente emanada desde los países que llevan la vanguardia en estas materias.

Es así como gracias a este benchmarking de sitios web, pude extraer características en común entre todos ellos como la norma esencial de adscribir los Principios FAIR; es decir, que los datos de investigación sean encontrables, accesibles, interoperables y reutilizables, hecho esencial en un repositorio de semejante naturaleza. Según Mark D. Wilkinson y otros autores, estos Principios “sirven para guiar a los productores y editores de datos mientras navegan por estos obstáculos, lo que ayuda a maximizar el valor agregado obtenido por la publicación académica digital formal contemporánea. Es importante destacar que nuestra intención es que los Principios se apliquen no solo a los ‘datos’ en el sentido convencional, sino también a los algoritmos, herramientas y flujos de trabajo que llevaron a esos datos. Todos los objetos de investigación digital académica -desde los datos hasta los procesos analíticos- se benefician de la aplicación de estos Principios, ya que todos los componentes del proceso de investigación deben estar disponibles para garantizar la transparencia, la reproducibilidad y la reutilización” (Wilkinson, 2016) .

Al respecto, es importante seguir avanzando y perfeccionando la legislación internacional (y nacional) para promover esta práctica de manera generalizada, encontrando acuerdos entre las partes que hoy se encuentran en conflicto: por un lado, los defensores de la CA y el AA; y, por otro, algunas editoriales y revistas especializadas que se cuestionan el tema de los derechos de propiedad intelectual.

¹⁷³ DataCite es una organización internacional sin fines de lucro cuyo objetivo es mejorar la cita de datos para establecer un acceso más fácil a través de Internet, teniendo como política la aceptación de datos de investigación como contribuciones legítimas y citables al registro académico. Sitio web: <https://datacite.org/>

¹⁷⁴ Directory of Open Access Journals (DOAJ). Sitio web: <https://doaj.org/>

Por otro lado, los sitios analizados presentaron otras características similares como hardware y software de calidad, con una gran capacidad de almacenamiento de datos; software en un idioma predeterminado (por defecto, inglés), el cual también permite traducción a otros idiomas (considerando que el usuario tenga una lengua materna diferente); interfaz adaptada al objetivo del software para permitir una búsqueda eficiente, a través de autor, título o fecha, por ejemplo, de una investigación; widgets para hacer más amigable la experiencia de navegación dentro del repositorio por parte de los usuarios, tanto de los investigadores como de los Bibliotecarios u otros usuarios; identificadores persistentes; metadatos y gestión de objetos digitales (datos, fotografías, infografías, estadísticas y grabaciones de audio o vídeo, por ejemplo).

Otro rasgo que destacó entre repositorios de datos de investigación se relacionó con la seguridad y confiabilidad, tanto en la gestión de almacenamiento de datos como en la diligencia de compartirlos. De esta manera, y debido a la existencia de una serie de estándares, se protege y promueve la cosecha de datos, así como la recopilación y la entrega de éstos dentro de un mismo sistema como hacia otros, garantizando la calidad y seguridad de los datos de investigación.

8. Aplicación a casos de estudio

En este capítulo analizo en detalle dos casos de estudio. Por un lado, el Repositorio Institucional del Centro de Ciencias de la Atmósfera, de la UNAM, y, por otro, un caso exitoso de aplicación de la plataforma Figshare, correspondiente a la Universidad Estatal de Iowa. En ambas experiencias abordo sus orígenes, desarrollo y expansión, características destacadas y visión del futuro.

8.1. Caso 1: REPOSITORIO INSTITUCIONAL DEL CENTRO DE CIENCIAS DE LA ATMÓSFERA (UNAM)

8.1.1. El origen

El académico e investigador de la UNAM, Agustín Fernández Eguiarte, lideró la creación de este repositorio que se inició como un proyecto en 2015¹⁷⁵. Posteriormente, ha habido nuevas etapas de implementación, con financiamiento de variada índole. Como se mencionó con anterioridad, este proyecto es uno de los casos más exitosos que se presentan en este documento, considerando su gestación, evolución, variables y futuro.

El Repositorio Institucional del Centro de Ciencias de la Atmósfera no nació bajo el alero de bibliotecas, sino como una respuesta a las necesidades de información del Centro de Ciencias de la Atmósfera en la UNAM. Me pareció una experiencia exitosa y, además, de sumo interés tener acceso a la opinión del fundador de este proyecto, razón por la cual establecí contacto digital con él, luego de invitar a dicha institución a responder la encuesta (aplicada también para el presente trabajo y cuyos detalles serán compartidos más adelante), recibiendo de vuelta de correo electrónico un comentario de Fernández Eguiarte, sintiéndome motivada para entablar comunicación con el investigador de manera más cercana. Fue así como, en mi opinión, lo anterior desembocó en una gratificante instancia de conversación¹⁷⁶ donde el investigador se refiere -entre otros temas- a la naturaleza de los datos de investigación en su disciplina y a los repositorios de datos, los cuales -afirma- deberían ser concebidos al interior de instituciones dedicadas a la investigación. Desde mi perspectiva, sus argumentos y postura fueron consideradas relevantes y dignas de ser compartidas en este capítulo para que las bibliotecas tengan acceso previamente a la mirada de un investigador que conoce el tema de los repositorios de datos en profundidad, liderando uno que, en 2019, fue distinguido con el galardón “Reconocimiento al Proyecto de Datos Abiertos 2019 en la Categoría Organismo Autónomo Federal en el marco de los Premios U-GOB ¹⁷⁷ 2019 Tecnología en Gobierno” (U-GOB. Tecnología en Gobierno, 2019).

¹⁷⁵ A. Fernández Eguiarte, comunicación personal, 23 de noviembre 2020, vía Zoom.

¹⁷⁶ A. Fernández Eguiarte, comunicación personal, 23 de noviembre 2020, vía Zoom.

¹⁷⁷ U-Gob es la iniciativa digital del Gobierno Federal de México. Sitio web: <https://u-gob.com/acerca-de-u-gob/>

Desde el ámbito de su dotación de personal, este repositorio se encuentra integrado por geomáticos, alumnos de la especialidad mencionada e ingenieros en computación. Se trata de un equipo que ha estado transitando con una dotación de entre cinco y hasta siete personas, con dedicación exclusiva para el proyecto y con otras incursiones en el ámbito investigativo, quehacer muy vinculado al uso de tecnologías de vanguardia, característica fundamental de la disciplina.

En la UNAM existe una Dirección General de Repositorios Universitarios (DGRU)¹⁷⁸, que tuvo su origen en el Instituto de Biología, responsable de crear un depósito de colecciones universitarias con información sobre especies biológicas y fotografías, entre otros datos, concibiendo un repositorio como la solución tecnológica adecuada al respecto. Paralelamente, se asentó el criterio de cómo clasificar estos objetos que, hasta dicho momento, eran diferentes a lo conocido: las publicaciones. El Instituto invitó a Fernández Eguiarte a colaborar, porque querían ampliar su espectro investigativo con otros datos de campo y porque él contaba con colecciones de datos de clima y de cambio climático. Así se gestó esta incipiente y sutil colaboración entre disciplinas de investigación de la Universidad. En dicha instancia se encontraban cuando, en 2015, CONACYT¹⁷⁹ emitió una convocatoria para crear repositorios institucionales. El investigador mexicano postuló a la primera convocatoria, resultando adjudicados tanto él como el resto del equipo con fondos para implementar un repositorio institucional de Ciencias de la Atmósfera.

En este contexto, inició un proceso de búsqueda y de análisis de las tecnologías más idóneas que permitieran desarrollar dicho proyecto. CONACYT emitió lineamientos para esta convocatoria basados en OpenAire, por lo que el incipiente grupo de proyecto debió adaptarse a las normas establecidas, aunque percibiendo que los lineamientos no se ajustaban a sus necesidades relacionadas con datos de clima y con cambio climático. Entonces, el desafío para este grupo de proyecto naciente fue adoptar tecnologías capaces de satisfacer sus intereses investigativos y que se ajustaran (cosechadas), simultáneamente, con los lineamientos exigidos por OpenAire en la convocatoria. Siendo la política de CONACYT propiciar la participación de bibliotecas y, considerando que en el Centro de Ciencias de la Atmósfera existía una biblioteca que entregaba servicios al mencionado centro y a Institutos de Ciencias de la Tierra, Fernández Eguiarte decidió contactar al encargado de biblioteca para invitarlo a sumarse al proyecto, con la idea de incorporar en éste algunas publicaciones.

En esta nueva colaboración, la Biblioteca hizo manifiesta su predilección por utilizar tecnología DSpace y el estándar Dublin Core, ambos recursos muy conocidos en este ambiente. No obstante,

¹⁷⁸ El objetivo principal de la DGRU es desarrollar e implementar tecnologías, metodologías y normatividad técnico-jurídica para los repositorios de la UNAM. Desde el inicio, ha sido la responsable de desarrollar, administrar, organizar, implementar y poner en funcionamiento el Repositorio Institucional de la Universidad. Asimismo, continuó con la operación y actualización del Portal de Datos Abiertos UNAM Colecciones Universitarias. Sitio web: <https://dgru.unam.mx/>

¹⁷⁹ Consejo Nacional de Ciencia y Tecnología (CONACYT). Sitio web: <https://www.conacyt.gob.mx/>

Fernández Eguiarte requería utilizar otras tecnologías del ámbito espacial: GeoNetwork¹⁸⁰ y el estándar ISO 19139¹⁸¹ para metadatos. Al no haber consenso entre la Biblioteca y el Centro de Investigación en cuanto al tipo de tecnología que se utilizaría para la implementación del repositorio, el investigador recibió una lista de publicaciones de parte de la Biblioteca de Ciencias de la Tierra para incluir en la nueva plataforma, lo cual originó una sección compuesta por artículos y tesis no sólo de maestría, sino también de doctorado. Esta propuesta de contenido, que incluía este componente de publicaciones y tesis, se presentó a CONACYT, que aprobó el proyecto por un año. Fernández Eguiarte continuó trabajando con su equipo de proyecto, desligándose de la vinculación con la Biblioteca, para enfocarse en la inclusión de nuevas variables al repositorio, con otros componentes de los datos geoespaciales, de ciencias de la atmósfera, de la Tierra y oceánicos, entre otros.

8.1.2. Datos de investigación: un gran tema para los investigadores

En palabras del investigador, el desarrollo de controles de calidad de los datos de investigación es la complicación más grande en este tipo de repositorios, a diferencia de los repositorios en bibliotecas, cuyas publicaciones han pasado procesos de evaluación por pares u otras instancias. Muchos datos científicos geoespaciales y de ciencias de la Tierra -manifiesta- no han sido filtrados con un control de calidad, siendo éste un aspecto fundamental para considerar en los datos científicos. Si esto no es abordado, se puede incurrir en la práctica de depositar muchos datos generados por diversas instituciones oficiales -sean nacionales o internacionales-, datos que desde su origen podrían no estar correctos, incluyendo errores graves como falta de calibración de un equipo, errores humanos e inconsistencias diversas. Dado lo anterior, los investigadores del área de climatología de esta disciplina se quejan, frecuentemente, por los errores al momento de graficar los datos, encontrando discontinuidades. Entonces, depositar los datos sin un adecuado proceso de control de calidad puede resultar una acción que genere resultados equivocados en las investigaciones, pues al registrar datos erróneos se puede desencadenar la realización de procesos errados con indeseadas consecuencias. Desde su mirada, esto no sucedería en las bibliotecas, pues siempre las publicaciones -con independencia de sus resultados- son sometidas a una evaluación y a un adecuado registro bibliográfico.

Los actuales datos de este repositorio reciben un exhaustivo control de calidad del cual son responsables los integrantes del equipo de proyecto. Para lograr lo anterior, han generado procesos descritos también en el metadato. Así, los usuarios puedan darse cuenta sobre lo que se hizo, lo que no o qué se encontró, gracias a esta característica incorporada en el proceso y a la metodología empleada para la estructuración de los metadatos. Por ejemplo, en el caso de los datos climáticos, éstos quedan visibles e integrados en los metadatos, tratamiento posible por la utilización de ISO 19139, a diferencia de Dublin Core, que no está diseñado para ello. El estándar que adoptaron tiene

¹⁸⁰ GeoNetwork es una aplicación para administrar recursos referenciados espacialmente. Proporciona potentes funciones de búsqueda y edición de metadatos, así como un visor de mapas web interactivo. Sitio web: <https://geonetwork-opensource.org/>

¹⁸¹ Norma ISO 19139 para metadatos geográficos. Sitio web: <https://www.iso.org/standard/32557.html>

campos, espacios para ingresar específicamente, porque está hecho para esto, ajustado a la naturaleza de los datos de la especialidad. Se trata de una característica muy importante para los investigadores de la disciplina, quienes disponen de un campo obligatorio para registrar la metodología y para describir cómo se hizo. Además, es una condición conocida y valorada por ellos.

En cuanto a la preservación del dato en el repositorio, éstos se van actualizando. Cabe señalar que los datos tuvieron un primer período de observaciones de 1902 a 2011 para, posteriormente, en una segunda etapa actualizarse para el período de 1902 a 2015, siendo que para ambas etapas se realizó un control de calidad. La primera etapa fue financiada por el Servicio Meteorológico Nacional de México y la segunda por el CENAPRED¹⁸².

Actualmente, están postulando un proyecto en colaboración con una agencia europea que tiene por objetivo realizar una nueva actualización, conservando los datos científicos previos de etapas anteriores que queden como referencia. Como estos datos climáticos se capturan de forma sistemática y continua, surgen sensores tanto satelitales como in situ nuevos, que resulta imperioso actualizar.

En relación a los obstáculos, Fernández Eguiarte concuerda que pueden surgir frente a un proyecto de esta naturaleza en aspectos como la búsqueda constante de financiamiento, la capacitación de personas que -luego- no se pueden mantener y ciertos aspectos culturales desde la perspectiva de comprender la especialización. Se debe tener presente que un especialista es un experto en su materia, pero lo cierto es que hay también especialidades relacionadas que, al momento de trabajar los datos de un repositorio, requieren de la competencia de un especialista en dicho dato, quien debe filtrarlos, procesarlos y depositarlos en una estructura de datos científicos (en este caso, el repositorio), según el nivel de especificidad requerido. Aquí la interacción entre especialistas en cada materia es crucial. He aquí, también, una de las causas para explicar el escaso desarrollo de repositorios de datos de investigación a nivel regional, a diferencia de lo que ocurre en Europa y en Estados Unidos.

Por lo anterior, afirma ser partidario de la creación de repositorios temáticos, por áreas del conocimiento capaces de contener datos. Como ejemplo mencionó la incursión del Instituto de Ciencias de la Atmósfera de la UNAM. Si bien están empezando, hay muchas áreas relacionadas que se podrían incorporar a dicha infraestructura tecnológica, iniciativas pendientes hasta el momento porque, para lograrlo, habría que trabajar con especialistas dedicados al estudio de las temáticas que se aborden y contar con el financiamiento para el desarrollo de la estructuración de la información y de los estándares que podrían utilizarse en un repositorio con las características mencionadas.

Actualmente, las bibliotecas están requiriendo de las revistas que publican artículos científicos la inclusión de los datos empleados en la investigación y, preferentemente, que éstos se encuentren

¹⁸² Centro Nacional de Prevención de Desastres (CENAPRED). Sitio web: <https://www.gob.mx/cenapred>

en un repositorio. Por eso, se trata de una práctica que ha ido ganando cada vez más terreno en algunas revistas científicas. Al momento de despachar los artículos, se solicita además el envío de datos y que éstos, de preferencia, hayan sido tomados de algún repositorio, de manera que queden ligados o vinculados a la publicación. En relación a dicha práctica, Fernández Eguiarte afirma que es favorable, porque estos datos científicos sustentan o dan apoyo a la investigación que se está publicando en una determinada revista científica.

Desde su perspectiva, considera que iniciativas como Zenodo¹⁸³ son favorables, ya que no corresponde sólo a un repositorio para datos científicos, sino que alberga además publicaciones científicas u otros elementos, sin dejar de integrar los datos científicos que sustentan la publicación.

Para el investigador existe una diferencia con respecto a los repositorios de publicaciones científicas que tienen datos asociados, los cuales se usaron para una investigación en específico. Esto no corresponde a una base de datos más general de alguna temática que alguien pueda consultar para generar otra investigación. Es diferente a datos satelitales, por ejemplo, de imágenes de otras fuentes que están en repositorios, pero que no fueron creados para asociar o vincular por el momento a una publicación en particular. Sin embargo, sí están estructurados con metadatos y con tecnología que un usuario puede consultar para su investigación, generando una publicación, citando estos datos o refiriéndolos, porque tomó de ahí esa información.

Esto que parece ser una sutil diferencia es de gran importancia para el ámbito de la investigación; por ende, se debe tener la mayor claridad posible al momento de implementar un repositorio para datos de investigación. Hay que definir si los datos corresponderán sólo a los empleados en una publicación o, por el contrario, se trata de un repositorio exclusivo para datos, enfatizando el investigador que habría que avanzar de temáticas particulares para -poco a poco- ir ampliando y conjugando con otras temáticas, es decir, no es recomendable iniciar con un repositorio multidisciplinario o generalista, sino por el contrario partir de la generación de repositorios temáticos, donde cohabiten -además- una serie de publicaciones junto con dichos datos correctamente registrados por disciplina. Su postura se sustenta en el hecho de que la generación y proliferación de datos científicos a nivel continental es enorme y el desarrollo de repositorios de datos científicos es aún incipiente, sumado a la escasez de políticas fuertes que incentiven esta materia y optimicen costos muy altos que implican la captura y procesamiento de la generación de datos científicos. Para Fernández Eguiarte, los datos de investigación que están en tecnologías implementadas por bibliotecas corresponden, preferentemente, a publicaciones que utilizaron esos datos para generar un artículo.

Por eso, su sugerencia es partir de forma progresiva, pensar que los datos hay que validarlos, así como la tecnología que se empleará. En la actualidad, se generan muchos datos sin estar exentos de un alto costo, en variados formatos y dispersos, duplicados sin ningún control de calidad; por

¹⁸³ Zenodo es un repositorio de Acceso Abierto y de propósito general, desarrollado por el programa europeo OpenAIRE y operado por la Organización Europea para la Investigación Nuclear (OEIN o CERN, en inglés). Sitio web: <https://zenodo.org/>

ende, plantea la necesidad de ir paso a paso, vinculado a especialistas y propiciando, además, una interrelación con informáticos cuya opinión será importante al momento de escoger una tecnología.

Comenta que en este repositorio hay datos de sensores satelitales de instituciones extranjeras (por ejemplo, de Francia) y, aunque el Instituto no tiene este tipo de sensores, toman dichos datos para procesarlos, adecuarlos y darles una utilidad regional. Aquellos investigadores internacionales que requieran depositar sus datos de manera directa, no pueden hasta sortear una revisión a cargo de este grupo de proyecto, quienes indican si están adecuadamente realizados con controles de calidad, cumpliendo todos los requisitos exigidos. Un aspecto importante de este repositorio es que coincide por completo con los lineamientos de INSPIRE¹⁸⁴, lo que también es un aspecto relevante de considerar al momento de implementar un repositorio: ajustarse a lineamientos de instituciones compatibles con el contenido temático de la tecnología empleada.

En relación a la licencia de los datos, en el caso de este repositorio, todos se encuentran con licencia Creative Commons, siendo datos abiertos y gratuitos, en bases temáticas alojadas en el repositorio. El Identificador Persistente (ID) de este repositorio es el metadato y está ligado a una cierta base de datos temática. En otras palabras, cada base de datos tiene un ID en su metadato.

Por otra parte, el investigador menciona que un aspecto crucial dentro de otras buenas prácticas corresponde a los Principios FAIR¹⁸⁵. En sus palabras “son fundamentales en estos momentos a nivel mundial, en materia de datos de investigación la tendencia de todos los países es adoptar estos principios”¹⁸⁶.

8.1.3. Detalle del Repositorio Institucional del Centro de Ciencias de la Atmósfera

8.1.3.1. Tecnologías empleadas

GeoNetwork Opensource	Aplicación informática de software libre y código abierto para la catalogación de recursos referenciados al espacio geográfico. Recomendada por iniciativas de infraestructura de datos científicos geoespaciales tales como: European Environment Agency (EEA), Infrastructure for Spatial Information in Europe (INSPIRE), World Meteorological Organization Information System (WIS Data), Observation Network for Earth USA (DataONE).
-----------------------	--

¹⁸⁴ INSPIRE es una plataforma de información integral para la comunidad de HEP, que comprende 8 bases de datos interconectadas sobre literatura, conferencias, instituciones, revistas, investigadores, experimentos, trabajos y datos. Sitio web: <https://inspirehep.net/help/knowledge-base/about-inspire/>

¹⁸⁵ Se trata de un conjunto de principios rectores, promovidos a nivel mundial, que deben cumplir los datos de investigación: Encontrable (Findable), Accesible (Accessible), Interoperable (Interoperable) y Reutilizable (Reusable). Sitio web: <https://www.nature.com/articles/sdata201618>

¹⁸⁶ A. Fernández Eguiarte, comunicación personal, 23 de noviembre 2020, vía Zoom.

ISO19139 Metadata Standard	Esquema internacional de metadatos diseñado para la descripción de datos geográficos derivado del estándar ISO19115 (ha sido homologado a Dublin Core y DataCite para cumplir con exigencias de financiamiento).
OAI PMH	Protocolo internacional para la interoperabilidad mediante el intercambio de metadatos.
GeoServer	Servidor para compartir e interoperar datos geoespaciales utilizando estándares abiertos del Open Geospatial Consortium (OGC).

8.1.3.2. Características destacadas

- Repositorio temático mixto, ya que integra datos científicos geoespaciales con publicaciones científicas.
- Cumple con los Principios FAIR: es encontrable, accesible, interoperable y reutilizable, en AA con licencia Creative Commons, su contenido permite generar nuevo conocimiento científico o ser de utilidad a quienes lo requieran.
- Proyecto LEARN CEPAL¹⁸⁷ al cual adhiere siendo un aporte en el proceso que ha iniciado la Comisión Económica para América Latina y El Caribe (CEPAL), en el marco del proyecto Leaders Activating Research Networks para la Gestión de Datos de investigación (GDI).
- Interfaz que permite la visualización cartográfica de datos geoespaciales.
- Propicia las Políticas para la CA y los Datos Científicos en América Latina de LA Referencia.

8.1.3.3. Interfaz web

Fernández Eguiarte comparte el home del Repositorio de la Agencia Europea de Medio Ambiente (AEMA o EEA, en inglés)¹⁸⁸, que utilizó GeoNetwork, entre otros recursos tecnológicos. Destaca la gran cantidad de temas que se ha agregado, convirtiéndose así en un avance significativo. El Repositorio Institucional del Centro de Ciencias de la Atmósfera es similar en su interfaz y en su estructura. Asimismo, se destaca la gran cantidad de bases de datos que componen este repositorio y el empleo de tecnología que permite su descarga de modo fácil. Comentó que en el Instituto decidieron realizar la transferencia de todo el contenido de la Agencia, tarea que les demandó quince minutos. El beneficio de esta cosecha es que lograron disponer de contenido que pueden publicar, al ser libre y abierto. El repositorio de la EEA también puede cosechar de sus temáticas, aunque son muy pocas aún. Destacó que este es un ejemplo de la interoperabilidad de los Principios

¹⁸⁷ El objetivo del proyecto LEARN es tomar como base la Hoja de Ruta LERU de Datos de Investigación, elaborada por la League of European Research Universities (LERU), como punto de partida para construir una infraestructura coordinada en Europa, Latinoamérica y el Caribe. Sitio web: <http://learn-rdm.eu/es/sobre-learn/resumen-del-proyecto/>

¹⁸⁸ European Environment Agency (EEA). Sitio web: <https://www.eea.europa.eu/es>

FAIR, enfatizando que no se trata de un repositorio de publicaciones, sino de un repositorio de datos científicos temáticos, utilizado para consultas, estudios y/o publicaciones, no estando en una biblioteca, sino en la EEA, con temáticas vinculadas y tanto con las estructuras tecnológicas como con los lineamientos de INSPIRE. Según el investigador, esta sería la forma en que hay que avanzar, tema por tema, siendo imposible abordar todas las temáticas de una vez, sino al contrario, de forma progresiva.

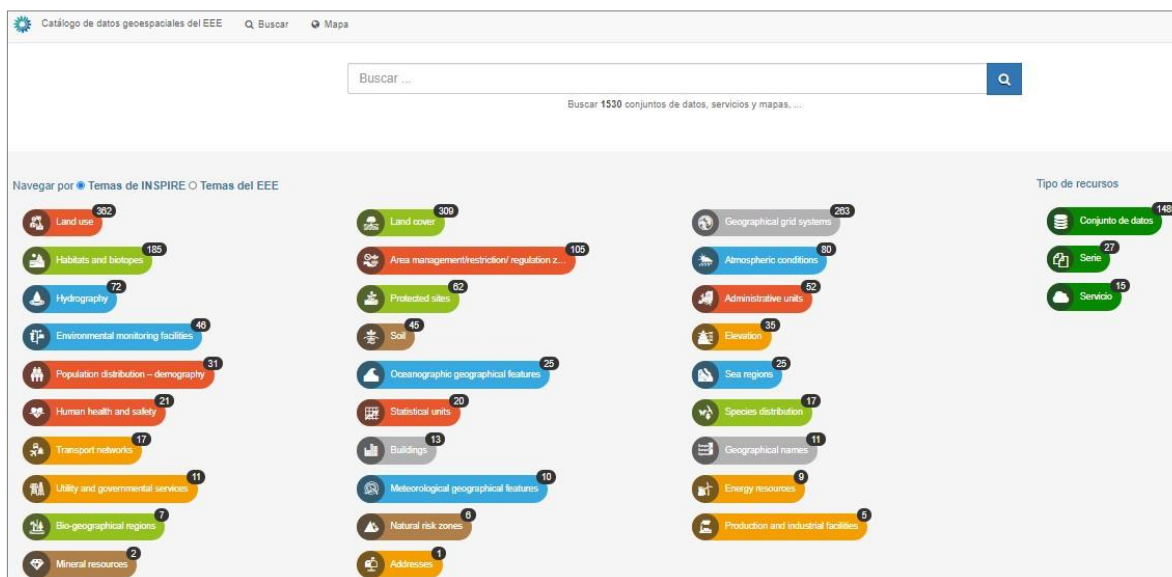


Figura 8-1: Home y temáticas del repositorio de la EEA

Fuente: <https://sdi.eea.europa.eu/catalogue/srv/eng/catalog.search#/home>





Figura 8-2: Home del Repositorio Institucional de Ciencias de la Atmósfera en la UNAM, con una interfaz inspirada en el repositorio de la EEA utilizando además tecnologías similares

Fuente: <http://ri.atmosfera.unam.mx:8586/geonetwork/srv/spa/catalog.search#/home>

8.1.4. Visión futura

Fernández Eguiarte considera que una biblioteca puede estar preparada para implementar un repositorio de publicaciones donde existan datos asociados, pero no puede -ni debería involucrarse- en el procesamiento de datos de investigación. Para crear una infraestructura de datos de investigación robusta que cumpla con todo lo abordado (por ejemplo, en materia de Ciencias de la Tierra), se requieren especialistas en la disciplina que deben estar involucrados directamente en el procesamiento de estos datos.

No obstante, considera que es un rol valiosísimo que las bibliotecas puedan disponer de repositorios y preservar los datos de publicaciones que -en la actualidad- son demandados por muchas editoriales o instituciones de financiamiento, cada vez con mayor insistencia. Por otra parte, reconoce y valora el rol formativo que tienen las bibliotecas, que pueden y deben entregar recomendaciones a los investigadores, tales como incentivar que éstos cuenten con su ORCID¹⁸⁹, lo cual les permitirá tanto dar visibilidad a los trabajos científicos como clasificarlos, tema muy desconocido aún y de alta necesidad.

¹⁸⁹ Open Researcher and Contribution ID (ORCID) proporciona un Identificador Digital Persistente. Sitio web: <https://orcid.org/>

8.2. Caso 2: FIGSHARE

Así como en el primer caso quise rescatar la experiencia del investigador Agustín Fernández Eguiarte respecto de los repositorios de datos de investigación -donde compartió sus puntos de vista sobre la participación de los investigadores y cómo éstos debieran interactuar y/o gestionar dicha tecnología-, como segundo caso quise rescatar la plataforma Figshare y la implementación de un repositorio de datos de investigación con este software (a modo de ejemplo, más adelante me referiré en particular al caso de la Universidad Estatal de Iowa).

Figshare fue concebido por un investigador y para investigadores. Con esto aludo a la génesis del proyecto, mientras Mark Hahnel concluía su Doctorado en Biología de Células Madre en el Colegio Imperial de Londres (Figshare, s.f.).

8.2.1. Origen y expansión

Esta tecnología para la implementación de repositorios de datos de investigación se concibió en 2011, aunque no fue sino hasta enero de 2012 que se lanzó al mercado, asociándose con ORCID en octubre de ese mismo año, lo cual permitió que -en adelante- este identificador se asignara automáticamente a quien lo requiriese. Luego, en 2014, pasó a las instituciones de modo más definitivo, dejando atrás las versiones de prueba beta. Hoy, tiene oficinas no sólo en Estados Unidos (Boston), sino también en el Reino Unido (Londres) y en Rumanía (Iasi).

Entre los logros que dan cuenta de su constante expansión en el mundo se encuentran: más de 26 millones de visitas a su página, más de 7.5 millones de descargas, más de 800 mil cargas por parte de los usuarios, más de 2 millones de artículos, más de 500 mil colecciones, más de 30 mil citas de contenido compartido y más de 5 mil proyectos. En cuanto a los usuarios, la mayoría corresponde a universidades o centros de investigación asociados a éstas, tanto en Estados Unidos, como en Europa, Australia, Nueva Zelanda, Asia y África¹⁹⁰. Llama la atención que no haya instituciones en México o en el resto de América Latina, aún declaradas en su web oficial, considerando la naturaleza misma del proyecto.

Según Patrick Splawa-Neyman, Especialista en Producto de Figshare: “Comenzó como un depósito de datos; un lugar para compartir y obtener crédito por los resultados que no encajan en una publicación estándar en papel. Esto se convirtió en un lugar para que los investigadores abrieran sus datos para respaldar la publicación de un artículo de revista. Desde entonces, la comunidad académica ha desarrollado una serie de aplicaciones novedosas para nuestra tecnología, que ha sido una de las cosas fortuitas que se derivaron de la construcción de una plataforma abierta. Creemos firmemente que Figshare es y seguirá siendo la plataforma de repositorio de datos más desarrollada” (Figshare, 2018).

¹⁹⁰ Sitio web: <https://figshare.com/>

8.2.2. Características destacadas

Figshare se ha definido como “el repositorio todo en uno”¹⁹¹, que incorpora perfectamente bien los Principios FAIR, ya que las investigaciones deben ser localizables, accesibles, interoperables y reutilizables, características también presentes en los resultados de investigaciones no tradicionales. Punto, éste último, donde resultan muy esclarecedoras las palabras del ya citado Patrick Splawa-Neyman: “Cuando hablo con los clientes, prefiero decirles que Figshare no es para datos, sino para archivos digitales. A menudo, el problema con el término datos es que cuando las personas piensan en datos, piensan en hojas de cálculo de Excel y archivos CSV de las disciplinas STEM¹⁹². Pero los datos pueden ser una grabación de audio o vídeo de un recital de poesía, pueden ser una encuesta, pueden ser un informe del Gobierno, un código o cualquiera de una multitud de tipos de archivos que son el producto de la actividad de investigación”. (Figshare, 2018)

Entre sus fortalezas se cuenta, también, que está diseñado como una herramienta que resulta de fácil gestión para investigadores, administradores y bibliotecarios, permitiendo -incluso- una interacción amigable con otros sistemas de una misma institución.

De hecho, los contenidos pueden publicarse en cualquier formato de archivo, asignando un DOI¹⁹³ a la marca institucional y siendo visibles en los principales motores de búsqueda y contextos académicos. Respecto de los documentos, consideran metadatos personalizados y específicos de la disciplina, estructurando las investigaciones según su naturaleza, departamento, laboratorio o tipo de archivo. Un aspecto importante -como indicó, también, el investigador de la UNAM, Agustín Fernández Eguiarte- es que la plataforma, antes de la publicación, se encarga de curar y revisar todos los contenidos¹⁹⁴, lo cual es un aspecto muy valorado por los investigadores, porque se relaciona con la calidad de los datos que se están compartiendo.

Tanto los desarrolladores de Figshare como el caso que se comentará más adelante, reconocen que la plataforma se adapta con flexibilidad al quehacer de los investigadores, automatizando la creación de registros al captar los datos de forma directa desde sus cuadernos de laboratorio electrónicos u otros instrumentos de investigación. Por otro lado, se reconoce la simplificación en el flujo de trabajo de investigación, abarcando los pasos de subir, gestionar, compartir y publicar los contenidos, los cuales pueden llegar a pesar hasta 5GB. “Aceptamos cualquier formato de archivo y

¹⁹¹ Sitio web: <https://figshare.com/about>

¹⁹² Science, Technology, Engineering and Mathematics (STEM). Es un término que sirve para agrupar estas cuatro áreas del conocimiento (Ciencia, Tecnología, Ingeniería y Matemáticas) en las que trabajan tanto científicos como ingenieros. Sitio web: https://es.wikipedia.org/wiki/Educaci3n_STEM

¹⁹³ Digital Object Identifier (DOI o Identificador de Objetos Digitales, en español) es una referencia única en línea a una página web con información sobre un objeto y cómo se puede acceder a él. Los objetos no requieren ser ni digitales para tener un DOI -ya que pueden físicos (como un libro)- ni estáticos, porque pueden cambiar con el tiempo. El registro DOI se actualiza si la dirección web o el contenido del objeto cambian de manera significativa, pero el enlace DOI siempre permanece igual. A diferencia de una dirección web URL, una referencia DOI nunca cambia, una vez creada. Sitio web: <https://doichile.cl/>

¹⁹⁴ Sitio web: <https://knowledge.figshare.com/institutions>

nuestro objetivo es obtener una vista previa de todos ellos en el navegador”¹⁹⁵, aseveran en el web oficial de Figshare.

Cada investigador dispone, además, de 20 GB de espacio privado sin costo, lo cual permite almacenar temporalmente su investigación de forma privada hasta que pase a AA. En este ámbito, se pueden compartir sus publicaciones, obtener un DOI por cada trabajo, citas bibliográficas adicionales para complementar los contenidos y facilidades para guardar los documentos en cualquier formato de archivo. Por otra parte, propicia las comunidades de investigadores que pueden compartir sus carpetas o enlaces con información o resultados de interés común, comportándose como una red social de comunicación científica, aspecto también muy valorado por dicho segmento.

Entre las especificaciones técnicas del software se destaca que Figshare adhiere a DataCite y a las mejores prácticas de DOI; es socio de lanzamiento de ORCID, con el propósito de mantener los perfiles automáticamente actualizados; puede actuar como backend¹⁹⁶ de almacenamiento para “Figshare for Institutions”, al encontrarse respaldado en DuraCloud¹⁹⁷; cuenta con un Comité de Ética de la Publicación, mediante el cual “nos adherimos a las directrices de COPE¹⁹⁸ sobre publicaciones académicas”; suscribe las directrices de UCISA¹⁹⁹ para estar en línea con los principales líderes de la educación superior; siguen los lineamientos de NISO²⁰⁰ para garantizar que su trabajo refleje los estándares adoptados a nivel mundial para los datos de investigación; y, por último, incorporan los postulados de CASRAI²⁰¹, en cuanto a reducir la carga administrativa de los investigadores, mejorando “la capacidad de inteligencia empresarial de las instituciones de investigación y los financiadores”²⁰².

¹⁹⁵ Sitio web: <https://figshare.com/about>

¹⁹⁶ Dentro del desarrollo web, el backend se encarga de todos los procesos necesarios para que la web funcione de forma correcta. Estos procesos o funciones no son visibles, pero tienen mucha importancia en el buen funcionamiento de un sitio web. Sitio web: <https://www.seoestudios.es/blog/que-es-backend-web/>

¹⁹⁷ DuraCloud es un proyecto de tecnología o software de código abierto para preservar y archivar contenido digital, siendo desarrollado y alojado como un SAAS por DuraSpace. Sitio web: <https://duraspace.org/duracloud/>

¹⁹⁸ Comité de Ética de las Publicaciones (COPE) se compromete a educar y apoyar a los editores, editores y aquellos involucrados en la ética de la publicación con el objetivo de llevar la cultura editorial hacia una donde las prácticas éticas se conviertan en una parte normal de la cultura editorial. Sitio web: <https://publicationethics.org/about/our-organisation>

¹⁹⁹ Universities and Colleges Information Systems Association (UCISA) se fundó en el Reino Unido en 1992 como un organismo profesional dirigido por miembros para profesionales digitales en educación. Abiertos e inclusivos, trabajamos juntos para utilizar nuestro conocimiento técnico colectivo y experiencia digital para ayudar a transformar la enseñanza, el aprendizaje y la investigación al respaldar la eficiencia operativa y una excelente experiencia estudiantil. Sitio web: <https://www.ucisa.ac.uk/About-us>

²⁰⁰ National Information Standards Organization (NISO u Organización Nacional de Estándares de Información, en español) se creó para identificar, desarrollar, mantener y publicar estándares técnicos para administrar la información. Es una organización sin fines de lucro. Sitio web: <https://www.niso.org/welcome-to-niso>

²⁰¹ Consortia Advancing Standards in Research Administration Information (CASRAI) es una iniciativa mundial sin fines de lucro fundada en Canadá, en 2006, para “fomentar activamente redes de colaboración entre las partes interesadas clave y diversas que recopilan información de los investigadores durante el ciclo de vida de la investigación y publicar pautas y recomendaciones estándar que permitan a los financiadores, proveedores e instituciones de investigación reducir la carga indebida”. Sitio web: <https://casrai.org/about/>

²⁰² Sitio web: <https://figshare.com/about>

En resumen, Figshare permite a los investigadores y a los usuarios recuentos de citas, control de versiones, opciones de almacenamiento personalizadas, funcionalidad colaborativa, actualizaciones continuas de productos, visualización de archivos, registros de metadatos y metadatos personalizados, entre otros. (Figshare, 2018) De esta manera, Figshare se ha convertido en un software reconocido en este ámbito y con un posicionamiento creciente.

8.2.3. Usuarios

Figshare se ha impuesto, paulatinamente, como el software preferido de las instituciones académicas gracias a sus postulados fundacionales que implican “gestionar, difundir y medir el impacto público de todos los resultados de su investigación; aumentar la visibilidad de su contenido, mientras brinda un mejor servicio a sus usuarios finales y autores; hacer que los resultados de su investigación estén disponibles de manera que se pueda citar, compartir y descubrir”²⁰³.

Entre las de mayor renombre se encuentran: Universidad de Melbourne (Australia), Universidad de Oxford (Reino Unido), Universidad de Auckland (Nueva Zelanda), Agencia de Protección Ambiental de los Estados Unidos (EPA, en inglés), Universidad Estatal de Iowa (Estados Unidos), Universidad de Ciudad del Cabo (Sudáfrica), Universidad de Estocolmo (Suecia), Universidad de Carnegie Mellon (Estados Unidos), Universidad de Ámsterdam (Países Bajos), Universidad de Arizona (Estados Unidos), Universidad de Administración de Singapur (Singapur) y 4TU.ResearchData (Países Bajos)²⁰⁴.

Al respecto, y en el contexto de la aplicación de la encuesta (enviada el 14 de octubre pasado y cuyos resultados serán compartidos más adelante), obtuve colaboración a través de correo electrónico de Andrew Mckenna-Foster²⁰⁵, Especialista en Producto de Figshare, quien realizó contribuciones a las preguntas relacionadas con el software propiamente tal. En un segundo contacto²⁰⁶, me compartió algunos ejemplos exitosos de repositorios de datos de investigación de instituciones que utilizan este software, quienes de acuerdo con sus necesidades pueden personalizar tanto la interfaz web como la administración de la misma. Destacó los siguientes casos: Universidad Estatal de Iowa (Estados Unidos), Universidad de Adelaida (Australia), Universidad de Lincoln (Nueva Zelanda), Universidad de Hong Kong (China), Universidad de Sudáfrica (Sudáfrica), 4TU (Países Bajos) y Universidad de Sheffield (Reino Unido). Posteriormente, en un tercer contacto²⁰⁷, agregó como caso relevante a la Universidad de Loughborough (también en Reino Unido), junto con enviarme la referencia de “El estado de los datos abiertos 2020”.

²⁰³ Sitio web: <https://knowledge.figshare.com/institutions>

²⁰⁴ 4TU.ResearchData es un repositorio de datos para ciencia, ingeniería y diseño, ofreciendo servicios como la conservación, el intercambio, el acceso a largo plazo y la preservación de conjuntos de datos de investigación procedentes de todo el mundo. Asimismo, ofrece asesoría a los investigadores para que sus datos de investigación se enmarquen en los Principios FAIR. Sitio web: <https://data.4tu.nl/info/en/about/organisation/>

²⁰⁵ Mckenna-Foster, A. (29-10-2020). Re: Survey on Research Data Repository

²⁰⁶ Mckenna-Foster, A. (23-11-2020). Re: Figshare Platform - successful cases

²⁰⁷ Mckenna-Foster, A. (07-12-2020). Re: Figshare Platform - successful cases

8.2.4. Visión futura

Si de visión futura se trata, para desarrollar este contenido tuve acceso al reciente informe “El estado de los datos abiertos 2020”, publicado el 1 de diciembre de 2020 por Ciencia Digital²⁰⁸, siendo el quinto reporte correspondiente a una encuesta que -en su desarrollo, aplicación y evaluación- contó con el apoyo de Springer Nature. Fueron contactados más de cuatro mil investigadores, quienes entregaron sus opiniones respecto de varias plataformas de repositorios y se refirieron, además, a tendencias en materia de datos abiertos.

Dentro de las interesantes opiniones registradas, destaco las expresiones de Mark Hahnel²⁰⁹ por ser el pionero de Figshare y por estar presente, nuevamente, en la mencionada publicación. De él rescato extractos de sus intervenciones por el significativo aporte que entrega en esta materia. Es así como en la primera pregunta se consultó sobre las mejoras que, según los investigadores, deberían producirse en el ámbito de los datos abiertos en los próximos cinco años, ante lo cual respondió: “Los indicadores de mejores prácticas serían un gran paso adelante en la profesionalización del espacio de publicación de datos. Durante la última década, hemos visto un crecimiento explosivo de investigadores que ponen a disposición conjuntos de datos. En Figshare, vimos el efecto positivo de la intervención humana para mejorar los metadatos en los conjuntos de datos financiados por los NIH (Institutos Nacionales de Salud). Sin embargo, lo que no tenemos son estándares para filtrar contenido. Hay una cantidad creciente de indicadores para filtrar, como FAIR (localizables, accesibles, interoperables y reutilizables), sellos de datos o citas que demuestren la reutilización. Las mejoras y la estandarización, en toda la infraestructura de datos aquí, nos moverán de ‘Muéstrame datos relevantes para mi investigación’ a ‘Muéstrame datos que sean reutilizables y relevantes para mi investigación’” (Hahnel, 2020).

Como puede apreciarse, Hahnel apuntó hacia la existencia de indicadores de buenas prácticas como un avance para profesionalizar el ámbito de la publicación de datos de investigación. Ante un explosivo crecimiento de conjuntos de datos -por cierto, disponibles para quienes lo requieran- se hace necesaria la existencia de estándares que permitan filtrar contenidos relevantes, lo que repercutirá en tener acceso y poder reutilizar, cuando otros investigadores lo requieran, datos específicos para intereses investigativos puntuales. Dentro de estos indicadores para filtrar contenidos, mencionó FAIR.

Al ser consultado sobre quién podría tener un mayor impacto para impulsar un cambio social en la comunidad investigativa o en la academia, conducente a datos abiertos, aludió a la necesidad de un cambio cultural considerado por muchos como necesario. Aquí Hahnel resumió magistralmente la

²⁰⁸ Digital Science es una empresa creada en Reino Unido, en 2010, y que se encuentra orientada a inversiones estratégicas para empresas emergentes focalizadas en emprendimientos relativos al ciclo de vida de la investigación. Su portafolio incluye empresas tales como: Altmetric, CC Technology, Dimensions, Figshare, Gigantum, GRID, IFI Claims, Overleaf, ReadCube, Ripeta, Symplectic y Writefull. Sitio web: <https://www.digital-science.com/>

²⁰⁹ Sitio web: https://figshare.com/authors/Mark_Hahnel/96371

situación con la expresión “Muéstrame los incentivos y te mostraré el resultado”²¹⁰, identificando así dos acciones arraigadas en el comportamiento de los investigadores que requieren, primero, avanzar en sus carreras para lo cual deben publicar y, segundo, lograr subvenciones para ser reconocidos y obtener éxito. En su opinión, ambas acciones están estrechamente vigentes y continuarán siendo los principales incentivos, los cuales determinan la conducta de los investigadores que deben publicar cada vez más datos y que deben estar, también, cada vez más abiertos a los financiadores.

Cuando se requirió su opinión sobre qué infraestructura tecnológica estaría faltando actualmente y qué se necesita implementar, en el contexto de los repositorios temáticos y de la gran cantidad de repositorios generales o multidisciplinarios, Hahnel planteó que los conjuntos de datos deben publicarse en un repositorio específico del tema, dentro de un tema, coincidiendo así con la mirada del investigador mexicano Agustín Fernández Eguiarte, desarrollada en el caso anterior. Sin embargo, hasta el momento, la mayoría de los datos generados se depositan en repositorios generalistas, debido a la escasez de repositorios específicos de materia.

En opinión de Hahnel, los repositorios temáticos deberían mantener requisitos de metadatos estrictos, muchas veces sometidos a curatoría de especialistas en una materia, lo cual repercutiría favorablemente en la recuperación. Sin embargo, aún se carece de mecanismos que permitan medir la calidad de los conjuntos de datos y metadatos, así como de mecanismos que promuevan recompensas para los investigadores comprometidos con la investigación y con los datos abiertos. Con frecuencia muchos repositorios declaran estar en AA, aunque carecen de un adecuado nivel de auditoría, que permita que lo anterior se cumpla a cabalidad. Para Hahnel, se trata de un tema relevante y grave a la vez que tiene relación con las API de los repositorios, las cuales deberían permitir la interoperabilidad, lo cual no siempre se cumple²¹¹.

En relación a las ganancias en materia de datos abiertos, Hahnel consideró que -tanto en el corto como en el mediano plazo- las ganancias son evidentes en el corto plazo, aludiendo a un estudio publicado en la revista Plos ONE, de abril de este año, de los autores McGillivray y otros²¹², quienes presentaron una estrecha asociación entre artículos, los cuales -declaran- enlazan a datos en un repositorio, redundando favorablemente en un mayor impacto de citas en promedio. De aquí se desprende que “si publica sus conjuntos de datos asociados con un artículo, tienen el potencial de

²¹⁰ “The State of the Open Data 2020. Digital Science Report”. McIntosh, Leslie et al. Data Science/Figshare, 2020. Págs. 9-10.

²¹¹ “The State of the Open Data 2020. Digital Science Report”. McIntosh, Leslie et al. Data Science/Figshare, 2020. Págs. 11-12.

²¹² “La ventaja de las citas de vincular publicaciones con datos de investigación”. Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K. y McGillivray, B. El artículo plantea la idea de que se están haciendo esfuerzos para que los resultados de la investigación sean abiertos y reproducibles, política que están adoptando las revistas de forma creciente, quienes alientan u obligan incluso a sus autores a proporcionar declaraciones de disponibilidad de datos. Sin embargo, aún no hay cifras definitivas respecto de qué porcentaje de estas declaraciones contienen realmente enlaces a datos bien formados (vía URL o identificador permanente) y de si existe un valor agregado en proporcionar dichos enlaces. Sitio web: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0230416>

obtener citas para su conjunto de datos más que en promedio”²¹³. En cambio, en el largo plazo, considera que lo que se pretende es hacer avanzar la investigación más rápido, eliminando los bloqueadores del acceso a la información. En este aspecto cita, además, el denominado “Cuarto Paradigma”²¹⁴.

Al ser consultado sobre los mecanismos que debiesen existir para que los datos gocen de un reconocimiento similar al de las publicaciones, Hahnel consideró que todos los productos generados por la academia deberían ser tratados de igual modo por todas las instancias que evalúan este tipo de productos. Así como el Índice H²¹⁵ es una métrica para medir el impacto en el ámbito de las publicaciones, no deberían existir razones para que los conjuntos de datos no sean también medidos, como por ejemplo con Google Académico.

8.2.5. Un caso exitoso de aplicación de Figshare: Universidad Estatal de Iowa

Como lo mencioné al inicio de este caso, tomé como ejemplo la experiencia -en este tópico- del repositorio de datos de investigación de la Universidad Estatal de Iowa (ISU, en inglés). Esta institución cuenta con una dilatada trayectoria académica (se fundó 1858), un alto número de estudiantes (alrededor de 34 mil 500) y está rankeada dentro de las más prestigiosas; así también está acreditada por la Comisión de Estudios Superiores en su país. Por otra parte, tiene suscritos convenios de intercambio estudiantil con otras instituciones no sólo en Estados Unidos, sino también en el extranjero, potenciando el conocimiento teórico y práctico de los estudiantes de pre y postgrado²¹⁶. Cabe recordar que su nombre oficial es Universidad Estatal de Ciencia y Tecnología de Iowa²¹⁷.

Entre las innovaciones que, en el último tiempo, ha implementado este centro de educación superior se encuentra el repositorio de datos de investigación “DataShare: The Open Data Repository of Iowa State University”²¹⁸, el cual fue un proyecto desarrollado en conjunto entre la Biblioteca y la Oficina del Vicepresidente de Investigación y Tecnología de la Información. Cabe puntualizar que DataShare²¹⁹ (también esta institución dispone de un repositorio institucional) es exclusivo para datos, está en AA y los datos han sido creados por académicos e investigadores de la Universidad y se pueden compartir y publicar.

²¹³ “The State of the Open Data 2020. Digital Science Report”. McIntosh, Leslie et al. Data Science/Figshare, 2020. Pág. 13.

²¹⁴ “El cuarto paradigma: descubrimiento científico intensivo en datos”. Hey, T. (autor), Tansley, S. (editor) y Tolle, K. (editora). Editorial Microsoft Research, 2009, 284 págs.

²¹⁵ Índice H fue desarrollado en 2005 por el físico Jorge E. Hirsch de la Universidad de California y su objetivo es cuantificar tanto la productividad científica real como el impacto científico aparente de un científico. Sitio web: <https://clinic-cloud.com/blog/que-es-el-indice-h-de-hirsch/>

²¹⁶ Iowa State University (ISU). Sitio web: <https://www.iastate.edu/>

²¹⁷ Iowa State University (ISU). Sitio web: <https://www.iastate.edu/>

²¹⁸ Sitio web: <https://iastate.figshare.com/>

²¹⁹ Sitio web: <https://datashare.icij.org/>

La Biblioteca es la encargada de gestionar este repositorio, razón por la cual me pareció -además- muy interesante de comentar. Esta administración tiene la ventaja y el beneficio de que tanto académicos como investigadores disponen de toda la asesoría y ayudas para la elaboración de un DMP o Plan de Gestión de Datos²²⁰.

Adicionalmente, y como parte del servicio, los usuarios disponen de una guía²²¹ muy completa, con varias secciones, las cuales contienen información referente a un módulo para aprender a redactar un plan, requisitos de las agencias y la posibilidad de consultar a través de correo electrónico, en caso de que un usuario, eventualmente, no encuentre aquí lo que necesita.

Por otra parte, se encuentra a disposición de la comunidad el acceso a DMPTool²²², una aplicación sin costo y de código abierto, la cual ofrece alrededor de cinco tipos de plantillas que permiten a los investigadores crear planes de gestión de datos para ser presentados en las instituciones o en las agencias financiadoras que así lo requieran.

En este caso, se puede apreciar cómo se manifiesta la relevancia del ciclo de vida de los datos científicos, donde componentes como la tecnología empleada -en este caso, los repositorios de datos- deben ser concebidos, considerando también la implementación de servicios, ya que aspectos como la asignación de metadatos (lo que conlleva describir, procesar y archivar), la necesidad de utilizar estándares de almacenamiento y el uso de licencias, entre otros, permitirán compartir y reutilizar de modo eficaz los datos de investigación. Sin embargo, como lo señalé, se requiere también como complemento la implementación de un servicio, el cual considere elementos desde la creación de planes de datos, entre otras tareas, las cuales repercutan favorablemente y promuevan las buenas prácticas que requiere esta materia.

También me ha parecido importante mencionar sobre este caso, la existencia de “formularios de divulgación de conjuntos de datos”, lo cual corresponde a una iniciativa del “Vicepresidente de Investigación, del Director de Información y de Biblioteca, quienes se comprometen a promover y mejorar el intercambio de datos de investigación producidos como resultado del trabajo realizado por los profesores, el personal y los estudiantes de ISU. Este formulario está diseñado para ayudar a recopilar ciertos atributos de los datos que se compartirán. Esta información se utilizará en el proceso de toma de decisiones para determinar si los datos se pueden compartir libremente”²²³. En el anexo 2 se encuentra el detalle de este formulario.

²²⁰ Sitio web: <https://open.lib.iastate.edu/open-data/data-management-plan>

²²¹ Sitio web: <https://instr.iastate.libguides.com/dmp>

²²² Sitio web: <https://dmptool.org/>

²²³ Sitio web: <https://app.smartsheet.com/b/form/de43938cc6f34868930c619d568e2dca>

Desde el punto de vista de la experiencia de usuario (UX²²⁴), me permití explorar exhaustivamente el sitio y apreciar su interfaz web, de la cual destaco su fácil accesibilidad y navegación a través de sus distintas secciones, las cuales comparto a continuación.

8.2.5.1. Interfaz web

En primer lugar, se despliega una pantalla general (Home) para acceder al repositorio de datos con cuatro secciones para la navegación: **Todas** / **Categorías** / **Grupos** / **Buscar**.



Figura 8-3: Header del home principal

Fuente: <https://iastate.figshare.com/>

Por defecto, la pantalla se despliega en la modalidad **Todas**, agrupando las publicaciones por relevancia en orden descendente (de la más a la menos reciente). Pudiendo, además, optar a los criterios de ordenar por primera cita online o por lista.

De esta forma, se despliegan horizontalmente cuatro conjuntos de datos por línea hasta el término de la sección, ordenados de izquierda a derecha según criterio de fecha de publicación (desde el más reciente, ubicado en el extremo izquierdo), cada uno de los cuales contiene -a su vez- cuatro entradas, las que permiten el acceso al contenido albergado en su interior. Dicho contenido equivale a los conjuntos de datos de cada uno, registrado con los metadatos específicos. A cada conjunto se puede acceder por una imagen relacionada con el conjunto de datos, por el título, por la referencia bibliográfica y/o por el investigador/creador.

²²⁴ Experiencia de Usuario (UX viene de User Experience, en inglés): es un concepto utilizado en Internet por medio del cual se alude a la percepción del usuario en su interacción con los sitios webs y todo lo que esto implica, como es el hardware, software, diseño gráfico, funcionalidad y accesibilidad, entre otros, teniendo como objetivo la “experiencia ideal” con respecto a dicho servicio o producto. Sitio web:

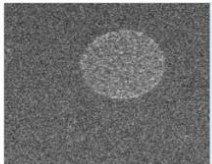
<https://www.uxmonkeys.com/blog/que-es-el-diseno-de-experiencia-de-usuario/>

Descubra la investigación de la Universidad Estatal de Iowa


+ Seguir

TODAS CATEGORÍAS GRUPOS BUSCAR

ordenar por: Relevancia



Transición de bobina a globular inducida por sal y ensamblaj...

Conjunto de datos publicado el 18.11.2020

Tanya Prozorov



Datos de selección de combustible y metabolismo celular...

Conjunto de datos publicado el 13.11.2020

Kaitlyn Holden



Datos de secuenciación de alto rendimiento (HTS) d...

Conjunto de datos publicado el 28.10.2020

Karthik Murugan



Efectos de los cultivos de cobertura a largo plazo en los...

Conjunto de datos publicado el 26.10.2020

Virginia Nichols

Volver arriba

Mostrar pie de página

Figura 8-4: Home del repositorio de datos de la Universidad de Iowa

Fuente: <https://iastate.figshare.com/>

Luego, ya en el conjunto de datos propiamente tal, se despliega el detalle del registro con los metadatos que identifican a ese conjunto de datos en particular, pudiendo encontrarse: figuras, título, detalle de investigadores con probabilidades de información ligada, una descripción del conjunto de datos, fondos de financiamiento, métricas de visita a este conjunto, posibilidades de compartir en redes sociales, categoría, palabras clave, tipo de licencia y opciones para exportar.

124

Datos de selección de combustible y metabolismo celular versus animal completo en la culebra de liga terrestre occidental (*Thamnophis elegans*)

Citar

Descargar todo (84,81 kB)

Compartir

Empotrar

+ Recoger

14 puntos de vista

13 descargas

0 citas

Conjunto De Datos publicado el 13.11.2020, 12:44 por [Kaitlyn Holden](#) , Ashley Hedrick , [Anne Bronikowski](#) , [Steven Hall](#) , Eric J. Gangloff

Estos datos incluyen información sobre culebras de liga terrestres occidentales (*Thamnophis elegans*), nacidas en cautiverio en el otoño de 2010 de hembras preñadas capturadas en la naturaleza. Se utilizaron animales para estudiar la relación entre la tasa metabólica de todo el animal y las tasas celulares de consumo de oxígeno y la principal fuente de combustible utilizada para apoyar el metabolismo, en un rango de temperaturas ambientalmente relevantes (24, 28, 32, 36°C).

FONDOS

Respuestas fisiológicas, genéticas y demográficas integradas al cambio de hábitat a largo plazo

Dirección de Ciencias Biológicas



CATEGORÍAS

- Fisiología animal - celular
- Metabolismo celular
- Ecología fisiológica animal

PALABRAS CLAVE

Figura 8-5: Ejemplo de registro de un conjunto de datos

Fuente: <https://iastate.figshare.com/>

La siguiente sección corresponde a **Categorías** desde donde se accede a 14 áreas: Ciencias Biológicas; Ciencias Agrícolas y Veterinarias; Matemáticas; Ciencias de la Tierra y del Medio Ambiente; Química; Ingeniería; Física; Diseño y entorno construido; Tecnología; Ciencias de la Información y la Computación; Psicología; Ciencias Sociales; Ciencias de la Salud y Comercio, Gestión, Turismo y Servicios. Para acceder a los contenidos, se utiliza el mismo criterio descrito anteriormente. Si la idea es realizar una búsqueda más específica, sólo hay que seleccionar una de las 14 categorías y, automáticamente, se despliegan subcategorías, condición que favorece la recuperación de datos al momento de la búsqueda.

Descubra la investigación de la **Universidad Estatal de Iowa**  [+ Seguir](#)

[TODAS](#) [CATEGORÍAS](#) [GRUPOS](#) [BUSCAR Q](#)

Ciencias Biológicas 

Data Summary

Abstract

Keywords



CONJUNTO DE DATOS



CONJUNTO DE DATOS



CONJUNTO DE DATOS

Datos de selección de combustible y metabolismo celular...

Conjunto de datos publicado el 13.11.2020

[Kaitlyn Holden](#) 

Datos de secuenciación de alto rendimiento (HTS) d...

Conjunto de datos publicado el 28.10.2020

[Karthik Murugan](#) 

Datos para: Franjas de vegetación de pradera colocadas dentro de...

Conjunto de datos publicado el 07.10.2020

[Matthew Liebman](#) 

Mejorar los resultados de las restauraciones de praderas en la ba...

Conjunto de datos publicado el 05.10.2020

[Lydia Inglés](#) 

 [Volver arriba](#)

Figura 8-6: Ejemplo de una categoría

Fuente: <https://iastate.figshare.com/category>

 [Vistazo](#) [Iniciar sesión](#)

Ciencias Biológicas  [+ Seguir](#)

[Biología Computacional](#)
[Microbiología](#)
[Genética](#)
[Biología Molecular](#)
[Metabolismo celular](#)
[Farmacología](#)

[Proteómica e interacciones intermoleculares \(excl. Proteómica médica\)](#)
[Biología evolucionaria](#)

[Bioquímica y biología celular no clasificadas en otra parte](#)
[Ecología comunitaria \(excluida la ecología de especies invasoras\)](#)

[Fisiología ecológica](#)
[Ecología de agua dulce](#)
[Ecología de la población](#)
[Ingeniería biológica](#)

[Expresión de genes \(incluido Microarray y otros enfoques que abarcan todo el genoma\)](#)
[Estructura y regulación del genoma](#)

[Genómica](#)
[Genética cuantitativa \(incluida la genética de mapas de rasgos y enfermedades\)](#)
[Micología](#)

[Fisiología animal - celular](#)
[Patología de planta](#)
[Biología celular y molecular animal](#)
[Ecología fisiológica animal](#)

[Estructura y función animal](#)
[Bioinformática](#)

[Todas](#) [BUSCAR Q](#)

Figura 8-7: Ejemplo de una categorías y subcategorías asociadas

Fuente: <https://iastate.figshare.com/category>

En resumen, se destaca que este repositorio asigna una relevancia a la utilización de metadatos, lo más específicos posible, tema también comentado por el investigador Agustín Fernández Eguiarte. Esta característica permite valorar a la especialidad, asignándole un rol destacado al momento de tratar los datos para su posterior recuperación y/o reutilización.

En relación a la sección llamada **Grupos**, se encuentran ocho en total, distribuidos de la siguiente manera: Agricultura, Arte y Diseño, Ciencias Biológicas, Negocio, Ingeniería, Ciencias Físicas y Matemáticas, Ciencias Sociales y Medicina Veterinaria.



Figura 8-8: Vista de grupos existentes

Fuente: <https://iastate.figshare.com/groups>

Al seleccionar cualquiera de ellos, se despliega una pantalla con las publicaciones agrupadas bajo la etiqueta correspondiente y se puede seguir el mismo patrón de búsqueda que en las secciones descritas anteriormente.

Por último, se presenta la sección **Buscar**, con una pantalla similar a la sección **Todas**, aunque presenta algunas diferencias, ya que aquí se incorporan una serie de filtros los que permiten hacer más específica la búsqueda y/o recuperación de datos. Estos filtros corresponden:

- Tipo de contenido:
 1. Artículo
 2. Colección
- Fecha de publicación:
 1. Fecha de publicación:
 - El mes pasado:
Desde: 09/11/2020
A: 09/12/2020
 - El año pasado:
Desde: 10/12/2019
A: 09/12/2020
 2. Primera cita online
 - El mes pasado:
Desde: 09/11/2020
A: 09/12/2020
 - El año pasado:
Desde: 10/12/2019
A: 09/12/2020
- Licencia:
 3. CC BY 4.0
 4. CC0
 5. MIT
- Tipo de artículo:
 1. Conjunto de datos
 2. Figura
 3. Medios de comunicación
- Grupo: muestra las 5 primeras opciones, permitiendo desplegar la lista completa de disciplinas al costado derecho de esta misma columna, donde también se indica entre paréntesis el número de artículos incluidos en ella.
- Categoría: muestra las 5 primeras opciones, permitiendo desplegar la lista completa de disciplinas al costado derecho de esta misma columna, donde también se indica entre paréntesis el número de artículos incluidos en ella.

De estos filtros, puedo añadir que el relacionado con las licencias, ofrece entre tres tipos (mencionadas también en el anexo 1). La licencia MIT fue recientemente agregada y corresponde a una licencia desarrollada por este Instituto como una licencia de software libre, permisiva, lo que se traduce en que prácticamente no impone restricciones respecto de la reutilización de los datos. Además, esta licencia es compatible con otras licencias del tipo copyleft.

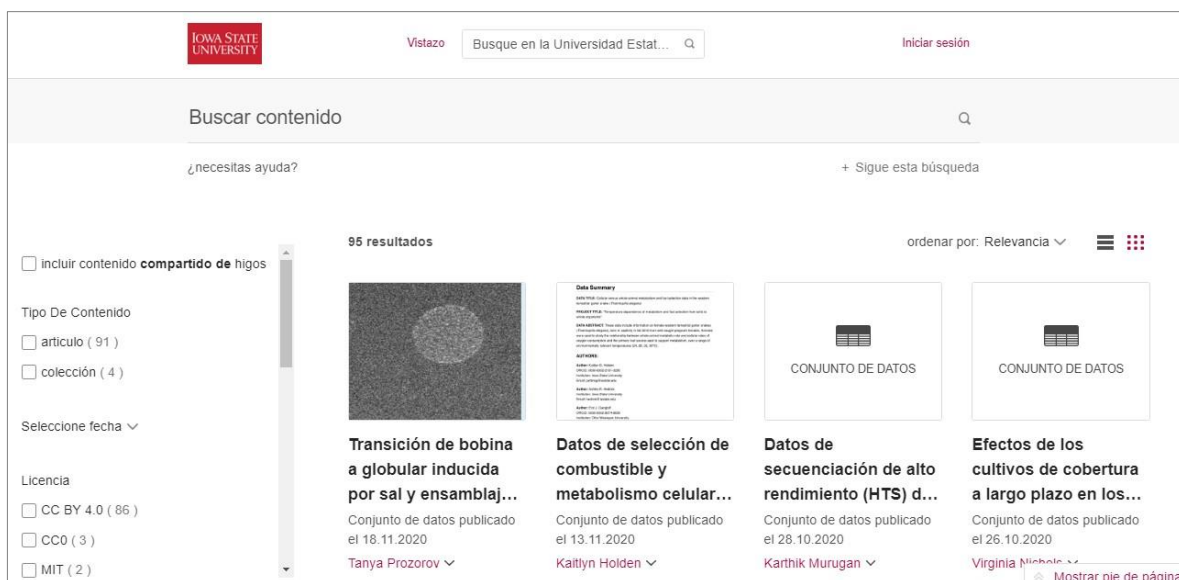


Figura 8-9: Ejemplo de la sección Buscar

Fuente: <https://iastate.figshare.com/search>

Como otra buena práctica o recomendación, este repositorio de datos incorpora, además, la puntuación y citación de Almetric²²⁵, como se presenta en la siguiente figura:

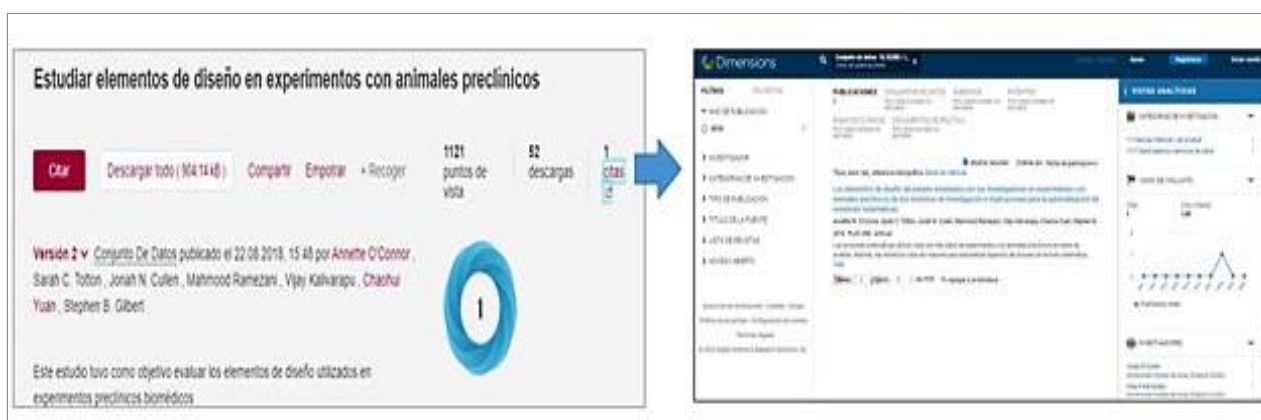


Figura 8-10: Detalle de la puntuación, citación en Almetric

Fuente:

<https://iastate.figshare.com/articles/dataset/Study design elements in preclinical animal experiments/6531224> y

https://app.dimensions.ai/discover/publication?and_subset_figshare_doi=10.25380%2Fiastate.6531224

Finalmente, destaco que DataShare está registrado en el directorio re3data.org y, también, en FAIR.Sharing.org²²⁶.

²²⁵ Es un conjunto de métricas que permite rastrear una variedad de fuentes para capturar y cotejar las conversaciones sobre contenido académico que a diario ocurren en línea. Sitio web: <https://www.altmetric.com/>

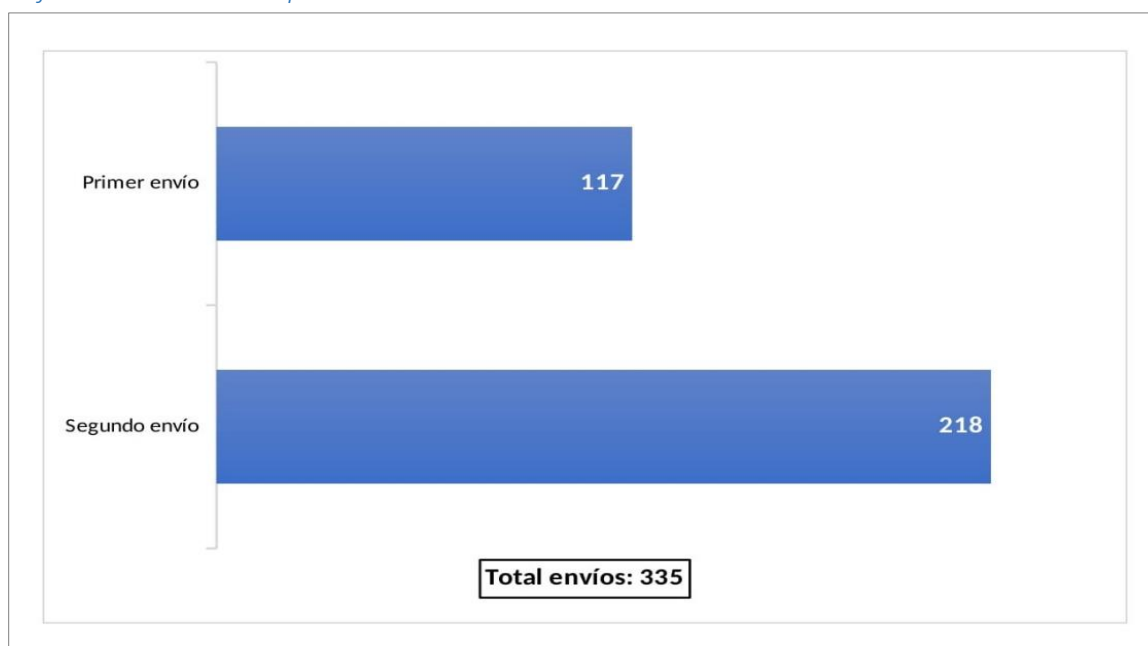
²²⁶ Sitio web: <https://fairsharing.org/> Ver también: <https://fairsharing.org/FAIRsharing.yOKiQs>

9. Resultados y análisis

En este capítulo, analizo en detalle los resultados obtenidos de una encuesta titulada **Repositorios de Datos de Investigación en Biblioteca**, la cual, según indagaciones realizadas, sería la primera vez que se aplica en nuestra área. El objetivo de este instrumento fue conocer las experiencias de diversas instituciones en el mundo, desde el software empleado, hasta inconvenientes y recomendaciones para implementar o mejorar la tecnología existente.

Con un total de quince sentencias en español e inglés, consideré preguntas con respuestas afirmativas/negativas, preguntas con opciones múltiples y preguntas con desarrollo. Se implementó a través de la plataforma LibWizard, de Springshare²²⁷. Se consideraron dos envíos que, en total, sumaron 335 instituciones con repositorios de datos de investigación de Acceso Abierto (AA).

Gráfico 9-1: Total envíos por encuesta

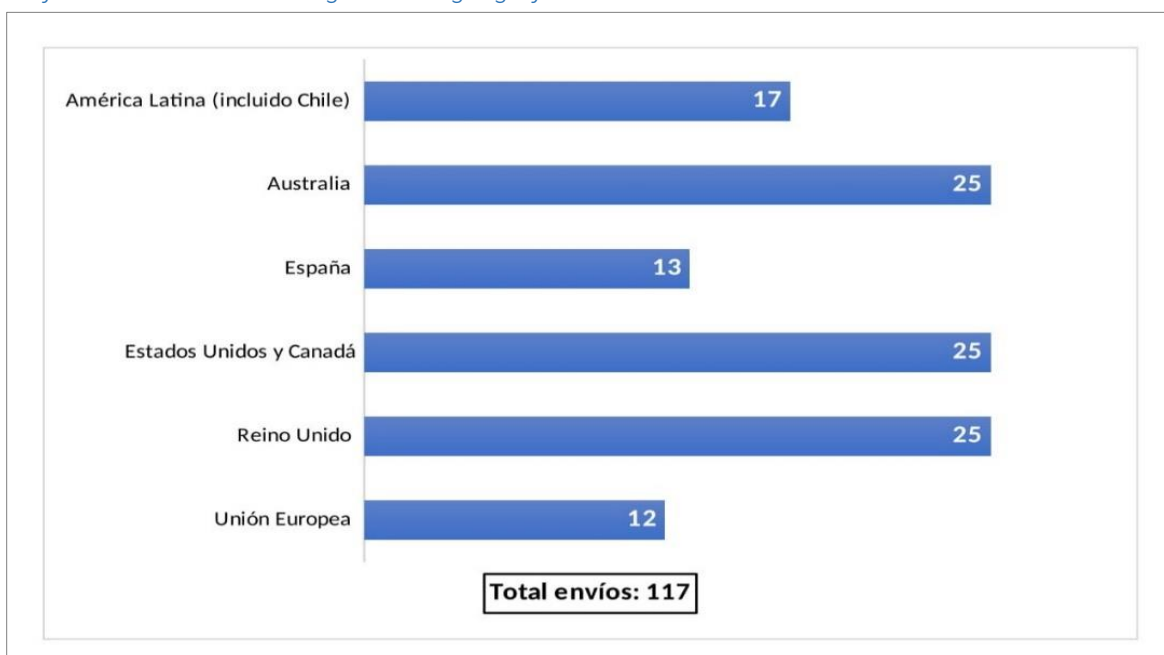


El primer envío se efectuó entre el 13 y 15 de octubre pasado, el cual incluyó un “gentle reminder” o recordatorio a las 117 instituciones requeridas, correspondientes al 34,93% del universo total de envíos. Dichos envíos se distribuyeron de la siguiente forma: 17 destinatarios en América Latina (incluido 2 en Chile), 25 en Australia, 13 en España, 25 entre Estados Unidos y Canadá, 25 en Reino Unido (separado de Europa, debido al Brexit)²²⁸ y 12 en la Unión Europea.

²²⁷ Sitio web: <https://spingshare.com/libwizard/>

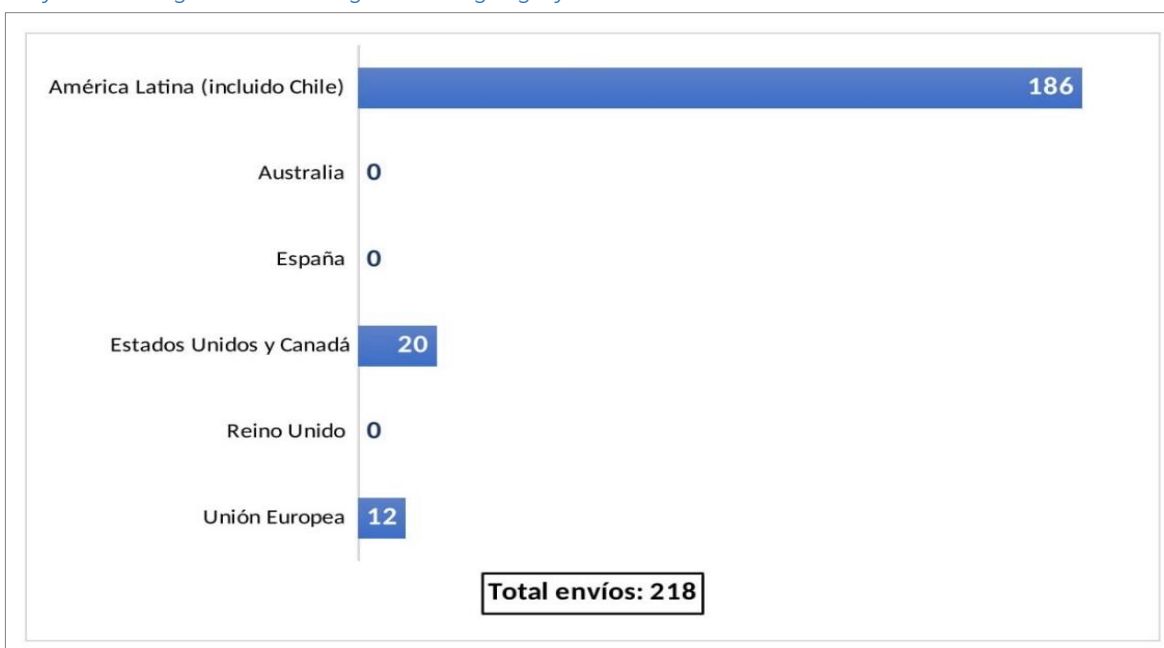
²²⁸ El Brexit es una abreviatura de las palabras inglesas Britain (Gran Bretaña) y exit (salida), siendo acuñado para aludir a la salida del Reino Unido de la Unión Europea, conglomerado que actualmente cuenta con 28 integrantes. Se unieron el 1 de enero de 1973 y, en junio de 2016, decidieron abandonarlo, lo que se materializó el 31 de enero de 2020. Hoy se estudia cuál es el tipo de relación que tendrá con ésta a partir del 31 de enero de 2021. Sitio web: <https://www.bbc.com/mundo/noticias-internacional-46521624>

Gráfico 9-2: Primer envío según zonas geográficas



En tanto, el segundo envío se realizó el 3 de noviembre del año en curso, dando por cerrada la encuesta el 12 del mismo mes. En esta ocasión, se consideraron 218 instituciones, diferentes al envío anterior, correspondiente al 67,07% del universo total de 335. Se distribuyeron de la siguiente forma: 186 en América Latina (incluido Chile), 20 en Estados Unidos y 12 en la Unión Europea.

Gráfico 9-3: Segundo envío según zonas geográficas



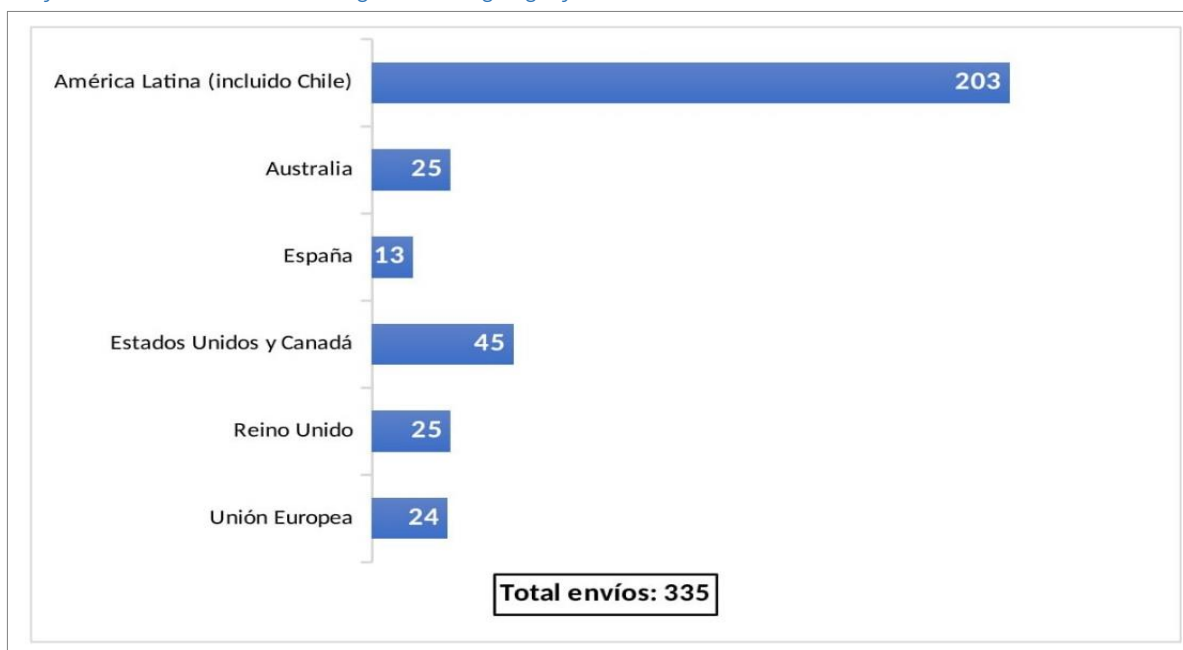
De esta forma, el total de las instituciones (335, equivalente al 100%) quedó distribuido por área geográfica de la siguiente manera:

- 203 en América Latina, equivalente al 60,60%
- 25 en Australia, equivalente al 7,46%
- 13 en España, equivalente al 3,88%
- 45 entre Estados Unidos y Canadá, equivalente al 13,43%
- 25 en Reino Unido, equivalente al 7,46%
- 24 en la Unión Europea, equivalente al 7,16%

En consecuencia, de acuerdo con los resultados de la encuesta, los países con mayor número de repositorios de datos de investigación de AA, agrupados en Europa y América del Norte, representaron el 31,93% del total de envíos.

Asimismo, en relación al envío de encuestas, se consideró relevante centrar éstas en instituciones en América Latina (casi un 61%), debido a que en esta zona geográfica -donde la literatura revisada reporta un incipiente desarrollo en esta materia- se espera, en consecuencia, que el análisis y los resultados presentados sean de utilidad para implementar o mejorar repositorios de datos de investigación.

Gráfico 9-4: Total de envíos según zonas geográficas



El criterio para la selección de las instituciones que fueron consideradas para este diagnóstico, se sustentó sobre la base de una exhaustiva búsqueda, en el ya mencionado, anteriormente, directorio

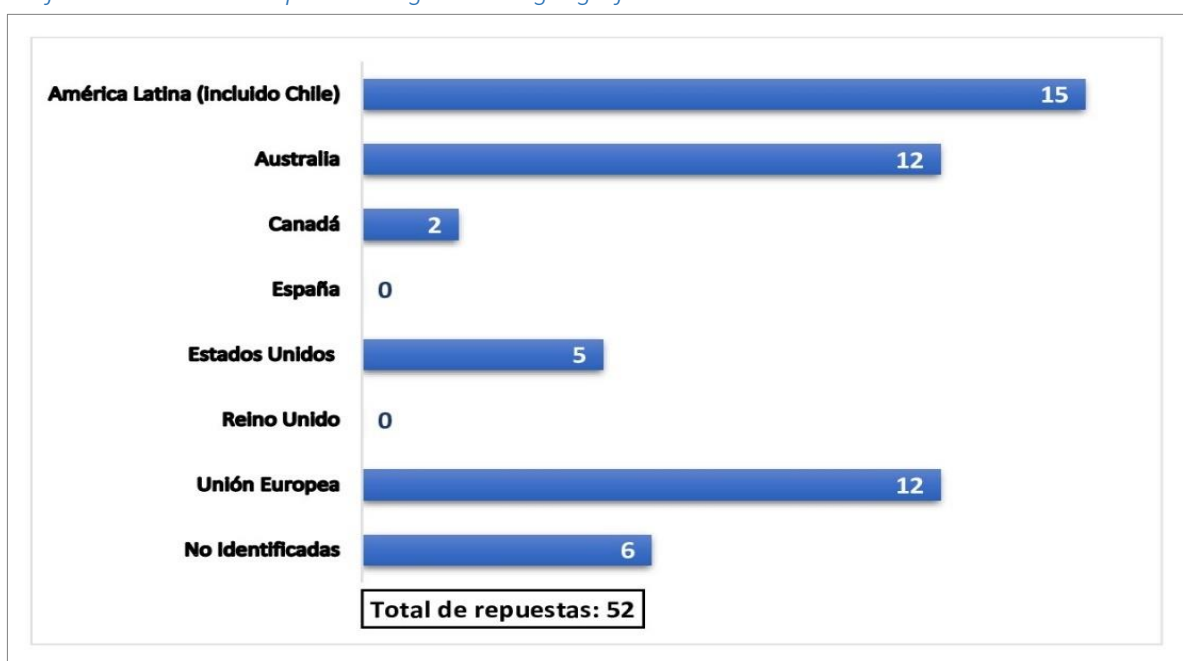
de repositorios de datos de investigación re3data.org²²⁹. En la primera selección se procuró elegir repositorios temáticos, en AA y, en lo posible, con alguna certificación. Asimismo, en la elaboración y metodología de la encuesta participaron un ingeniero experto en proyectos tecnológicos y un periodista especializado en comunicaciones institucionales. Este trabajo académico e interdisciplinar fortaleció y enriqueció la encuesta, convirtiéndola así en un instrumento más certero en la indagación de los datos y, asimismo, más expedito en cuanto a la administración y análisis propiamente tal. De esta manera, se acogió la sugerencia de la profesora guía, para quien esta instancia colaborativa generaría frutos respecto del estado de arte de las materias abordadas a lo largo de este trabajo.

En cuanto a los resultados, se obtuvieron 52 encuestas de los 335 envíos, lo cual representa una tasa de respuestas de 15,52% del total. Cabe hacer notar que las identidades de las instituciones que participaron en la encuesta se mantienen en absoluta reserva, de acuerdo a normativas nacionales e internacionales en esta materia, utilizando dicha información desde una perspectiva genérica que ilustró las tendencias sobre las materias encuestadas, poniendo de manifiesto semejanzas y/o diferencias sobre la realidad que actualmente caracteriza a los repositorios de datos de investigación en las áreas examinadas.

A continuación, se presentan gráficos de barra, extraídos desde el programa Excel de Microsoft, y gráficos de mesa, con los datos obtenidos de la plataforma LibWizard de Springshare, junto con algunas conclusiones, las cuales permiten un mejor análisis de los resultados obtenidos.

²²⁹ Registry of Research Data Repositories (re3data.org). Se trata de un servicio ofrecido por DataCite, organización sin fines de lucro, que proporciona identificadores persistentes para datos de investigación (Digital Object Identifier, DOI). En la actualidad, tiene como usuarios a más de 2 mil instituciones a nivel mundial. Sitio web: <https://blog.datacite.org/re3data-science-europe/>

Gráfico 9-5: Total de respuestas según zonas geográficas



En este gráfico se muestra el total de respuestas según zonas geográficas. Podemos concluir que si bien América Latina lideró en cantidad de instituciones contactadas para este análisis (203 en total, distribuidos en dos envíos a instituciones diferentes), lo cual repercutió favorablemente en la cantidad de encuestas contestadas (15 en total), esto representa sin embargo sólo un **7.4% de respuestas**, versus un **92.6 %** de instituciones que no contestaron. También se puede inferir que Australia, país que tiene un reconocido liderazgo en la temática, tuvo una participación significativa, ya que de las 25 instituciones consultadas (sólo hubo un envío), respondieron 12, cifra **equivalente a un 48%** (versus un **52%** que no contestaron). Esto convierte a Australia en el país quien tuvo la mayor tasa de respuestas a la consulta efectuada. Por otra parte, esta cifra fue muy similar a las respuestas de la Unión Europea, exceptuando que -en este caso- hubo dos envíos de la encuesta (en diferentes fechas), en cada caso contactando a 12 instituciones, lo que arroja un total de 24 envíos y 12 respuestas, equivalentes a un 50%.

9.1. Análisis de los resultados

A continuación, se examinan las preguntas una a una. El detalle de éstas se encuentra, además, completo en el anexo 3.

Pregunta 1:

¿Cuál es la dotación de personas que trabajan exclusivamente en el Repositorio de Datos de Investigación en su institución?

Con esta consulta se abrió la encuesta, donde se les preguntó a las instituciones acerca de la dotación de personas que trabajan en repositorios de datos de investigación, reflejando, desde el punto de vista de recursos humanos, una situación generalizada que da cuenta de un número más bien bajo destinado a cumplir con estas labores, ya que 41 instituciones (equivalentes al 78,85%) respondió tener entre **Menos de 5**. Muy por debajo, la siguiente opción más votada fue **Entre 5 y 10**, lo cual se traduce en un 9,62%. Cercano a este porcentaje se ubicó la opción **Más de 10**, englobando el 7,69% de las instituciones participantes. Cabe hacer notar que 1 institución respondió que su realidad no se ajustaba a ninguno de los criterios definidos en la pregunta (1,92%), mientras que otra se abstuvo de contestar (1,92%). Esto evidencia que, hasta ahora, una minoría dispone de un mayor número de personas para atender los requerimientos necesarios para la gestión de un repositorio de datos de investigación. Lo anterior podría relacionarse con lo incipiente aún del tema, factores presupuestarios y/o políticas en torno a los repositorios de datos de investigación, entre otros.

Gráfico 9-6: Dotación de personal



Gráfico 9-7: Dotación de personal

1. ¿Cuál es la dotación de personas que trabaja ...		
Respuesta	Número de envíos	% de envíos totales
Menor de 5 / Menos de 5	41	82,00%
Entre 5 y 10 / Entre 5 y 10	5	10,00%
Más de 10 / Más de 10	4	8,00%
Total	50	100,00%

Respuestas totales: 50

Pregunta 2:

¿Cuál es el nombre del software con el cual implementó su repositorio?

Si bien la pregunta anterior nos da luces acerca de la importancia que tiene el factor recurso humano en el desarrollo de los repositorios de datos de investigación, no menos relevante es el software que las instituciones eligen para implementarlos y administrarlos, punto abordado en la pregunta 2 de la encuesta y que arrojó interesantes resultados. De las 52 respuestas equivalentes al 100% de participación, 19 organizaciones (36,54%) declararon que el software implementado es **DSpace**, que se plantea como “el software de elección para organizaciones académicas, sin fines de lucro y comerciales que crean repositorios digitales abiertos”²³⁰, siendo utilizado en la actualidad por más de mil organizaciones en el mundo, entre las que se encuentran universidades, institutos de especialidad, academias, bibliotecas, bancos y ministerios gubernamentales²³¹, distribuidos principalmente en Europa y América del Norte.

En segundo lugar, y casi tres veces más atrás, se ubicó el software **Dataverse**²³², con 7 menciones (13,46%), y que, según su portal oficial, se define como “una aplicación web de código abierto para compartir, preservar, citar, explorar y analizar datos de investigación, facilitando la puesta a disposición de otros de los datos y le permite replicar el trabajo de otros con mayor facilidad”²³³. Este software se encuentra extendido de manera general en África, América, Asia, Europa y Oceanía. Recientemente, y en su política de mejoras constante, el sitio anunció dos medidas importantes como son 5GB de carga de almacenamiento adicionales para los administradores (08-10-2020) y un rediseño del conjunto de datos y archivos en integración con Google Cloud Archiver (09-11-2020)²³⁴. En tercer lugar, las preferencias arrojaron un empate entre las instituciones consultadas, ya que 3 optaron por **Fedora** (5,77%), mientras que otras 3 por **Figshare** (5,77%).

El resto de las instituciones que contestaron la encuesta se inclinó por diversos softwares, algunos de los cuales gozan de reputación entre los expertos. Es importante destacar que los mismos softwares continuamente experimentan mejoras y actualizaciones en pro de un mejor rendimiento, al tiempo de hacerlos más amigables para los usuarios.

Dicha dispersión podría dar cuenta de lo incipiente que es el tema de la habilitación y desarrollo de los repositorios de datos de investigación en las diferentes zonas geográficas. Incluso, dentro de una misma zona, hay diferentes grados de avance que se relacionan con la envergadura de la institución, dotación de personal, recursos económicos disponibles y grado de tecnologización informática con que cuentan estas instituciones dentro de sus respectivas planificaciones estratégicas.

²³⁰ Sitio web: <https://duraspace.org/dspace/>

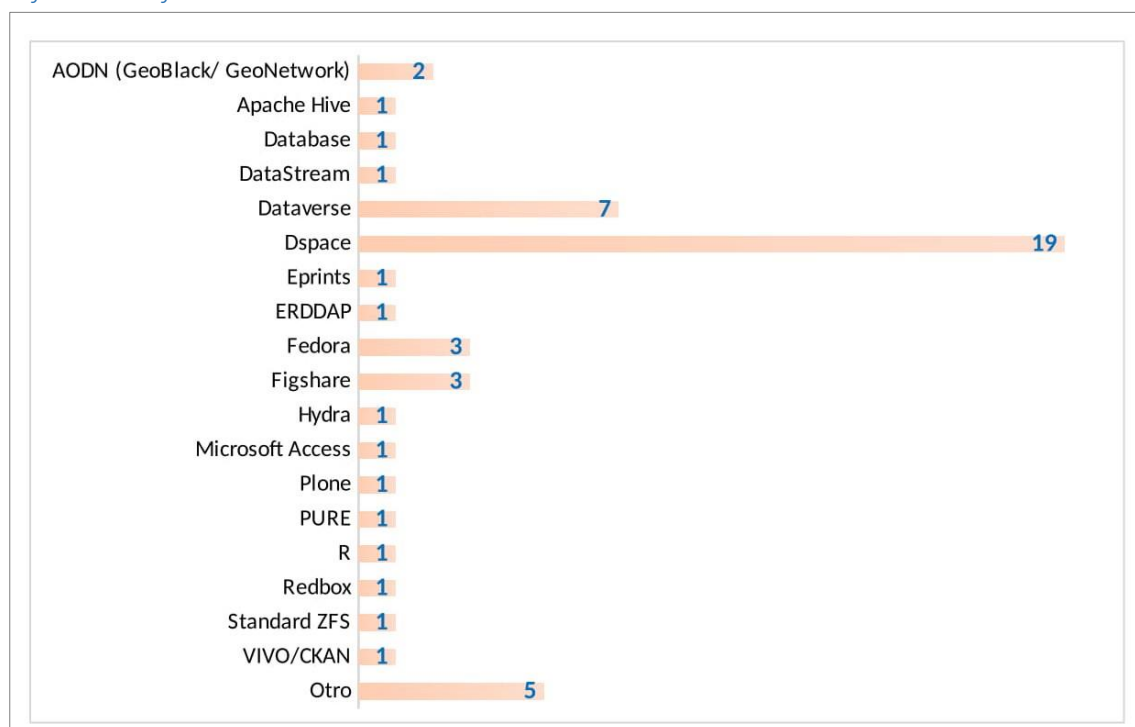
²³¹ Sitio web: <https://duraspace.org/registry/>

²³² Sitio web: <https://dataverse.org/>

²³³ Sitio web: <https://dataverse.org/about>

²³⁴ Sitio web: <https://www.iq.harvard.edu/roadmap-dataverse-project>

Gráfico 9-8: Softwares más usados



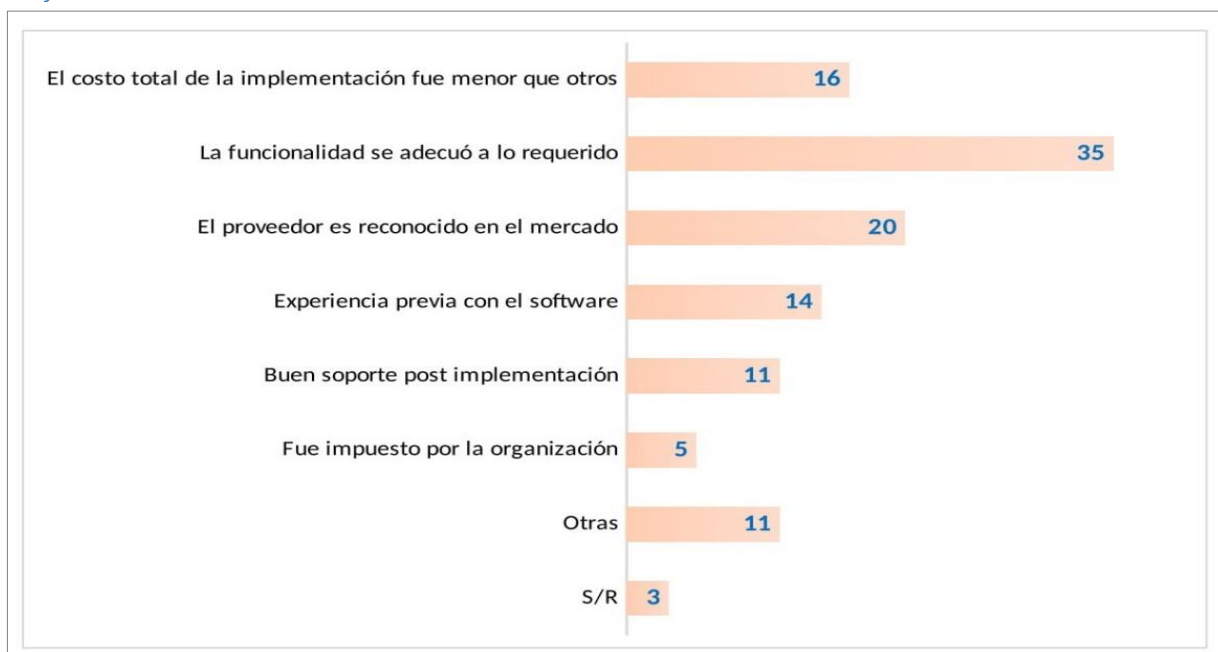
Pregunta 3:

¿Qué elemento fue clave en la selección del software (3 opciones como máximo)?

En línea con las respuestas a la Pregunta 2, claramente aquellos softwares que se encuentran mejor posicionados en el mercado, de lo cual dan cuenta las preferencias señaladas anteriormente, son los que reúnen los atributos más valorados por parte de las instituciones para implementar sus repositorios de datos de investigación. Entre éstos se cuentan **La funcionalidad se adecuó a lo requerido** (67,31%), el más votado de los 6 presentados en la encuesta, un aspecto esencial que demuestra que no sólo de la facilidad para agregar contenidos de los repositorios de datos, sino también de la accesibilidad a éstos por parte de los usuarios. Otro factor vinculado con la respuesta anterior se relaciona con **El proveedor es reconocido en el mercado** (38,46%), evidenciando su prestigio influye de manera decisiva en la elección del software, lo cual para las instituciones que respondieron se traduce en confianza hacia la marca sustentada en la garantía, credibilidad y profesionalismo brindados por ésta (este punto se vio como parte del capítulo anterior). No menor como variable determinante al momento de elegir un software es el presupuesto, razón por la cual en el tercer lugar aparece **El costo total de la implementación fue menor que otros** (30,77%), idealmente menor al de otras opciones en la licitación. Los dos últimos criterios decisivos para las instituciones fueron la **Experiencia previa con el software** (26,92%) y la entrega de un **Buen soporte post implementación** (21,15%), confirmando la tendencia que cuando un cliente (en este caso, una institución) tiene una buena experiencia en la compra de un bien o servicio a un proveedor (desarrolladores de softwares), probablemente continuará eligiendo en el futuro sus productos y servicios. De la mano de este último punto se sitúa el soporte entregado por el proveedor al cliente,

el cual debe ser expedito, eficiente y oportuno, sobre todo cuando se trata de repositorios de datos, donde se almacenan millones de datos con un crecimiento exponencial y deben estar accesibles a los usuarios en línea y tiempo real, evitando hacer colapsar el sistema. Aquí se manifiestan los Principios FAIR²³⁵.

Gráfico 9-9: Criterios de selección



²³⁵ Se trata de una serie de principios concisos y medibles que fueron desarrollados conjuntamente por la academia, la industria, las agencias de financiación y los editores académicos, teniendo como objetivo servir de guía para mejorar la reutilización de datos. Estos son: Encontrable (Findable), Accesible (Accessible), Interoperable (Interoperable) y Reutilizable (Reusable). Sitio web: <https://www.nature.com/articles/sdata201618>

Gráfico 9-10: Criterios de selección

3. ¿Por qué escogieron este Software? / Por qué...		
Respuesta	Número de envíos	% de envíos totales
El costo total de la implementación fue menor que otros / ¿El costo total de la implementación fue menor que los demás?	dieciséis	32,65%
La funcionalidad se adecuó a lo requerido / ¿La funcionalidad fue según lo solicitado?	35	71,43%
El proveedor es reconocido en el mercado / El proveedor es bien conocido en la industria.	20	40,82%
Experiencia previa con el software / Experiencia previa con el software	14	28,57%
Buen soporte post implementación / Buen soporte post implementación	11	22,45%
Fue impuesto por la organización / Fue impuesto por la organización	5	10,20%
Otro	11	22,45%
Total	49	100,00%

Respuestas totales: 49

Pregunta 4:

El repositorio de su institución, ¿es exclusivamente para datos?

Esta es una pregunta que apuntó directamente a la naturaleza de los repositorios de datos de investigación, los cuales, en su mayoría, se encuentran en instituciones académicas con el objetivo de estar disponibles para investigadores, bibliotecarios y administrativos en tiempo real, por medio de una operabilidad de fácil uso, y que pueden albergar no sólo datos, sino también otros contenidos, como expresa la plataforma Figshare, entre cuyos clientes se encuentran importantes universidades (Figshare, 2020). Desde esa perspectiva, se entiende que 28 instituciones (53,85%), equivalentes a poco más de la mitad de las respuestas, hayan declarado que sus repositorios son exclusivamente para datos. Como contrapartida, sólo 24 instituciones (46,15%) declaró que sus repositorios son para datos y para otros contenidos, como tesis, archivos fotográficos y/o audiovisuales. Esto evidencia la necesidad por parte de las instituciones de generar repositorios que brinden la mayor cantidad de información -y de calidad, en lo posible- no sólo para los investigadores, académicos y estudiantes, sino también para las personas que deseen acceder a dichos contenidos.

Gráfico 9-11: Repositorios de datos

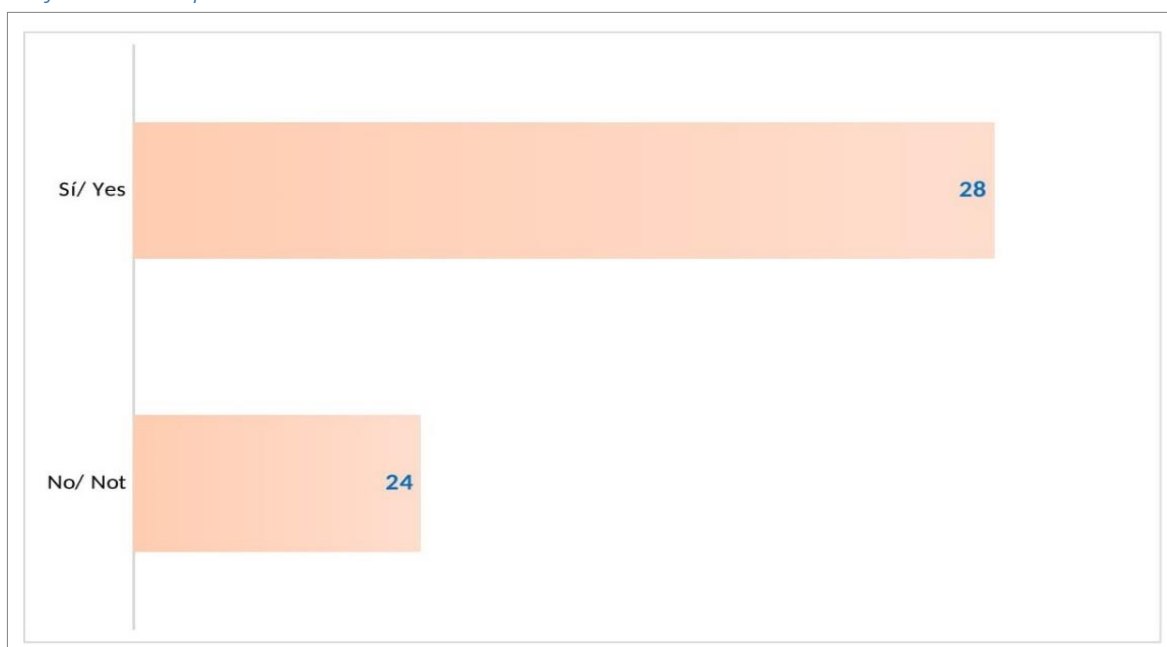


Gráfico 9-12: Repositorios de datos

4. El repositorio de su institución, ¿es exclus ...		
Respuesta	Número de envíos	% de envíos totales
Sí / Sí	28	53,85%
No no	24	46,15%
Total	52	100,00%

Respuestas totales: 52

Pregunta 5:

¿Tiene el repositorio de su institución una sección para datos de investigación, entre otras secciones?

Similar a los resultados mostrados en la pregunta anterior, en ésta también hubo 28 instituciones (53,85%) que declararon tener una sección para datos de investigación dentro de sus repositorios, cohabitando con otras secciones (generalmente tesis). Esta cifra puede englobar, probablemente, a las 24 instituciones que en la respuesta anterior declararon que sus repositorios no eran sólo de datos, más otras instituciones que pudiesen tener repositorios híbridos o mixtos, constituidos por una sección preferente de datos y otras secciones de una naturaleza diferente. Llama la atención, asimismo, el alto porcentaje de instituciones, 19 de 52 (correspondiente al 36,54%), que no entregó una respuesta frente a lo encuestado, impidiendo así hacer un levantamiento acerca de cómo se encuentran distribuidas las secciones dentro de sus repositorios.

Gráfico 9-13: Secciones para datos y otras



Gráfico 9-14: Secciones para datos y otras

5. ¿Tiene el repositorio de su institución una ...

Respuesta	Número de envíos	% de envíos totales
Sí / Sí	19	79,17%
No no	5	20,83%
Total	24	100,00%

Respuestas totales: 24

Pregunta 6:

¿Hace cuánto tiempo el repositorio de su organización fue implementado para datos de investigación?

Si tomamos como año de referencia 2020, las respuestas a esta pregunta nos estarían demostrando, al parecer, un repunte en la implementación de los repositorios de datos de investigación, luego de experimentar un estancamiento que sucedió a la creación de los primeros en el mundo, los cuales se generaron en respuesta a las necesidades de los centros de documentación de diversas instituciones. Sumando a lo anterior, habría que considerar -además- la transversalidad cada vez más fuerte, la proliferación del AA y la gestación tanto como maduración de políticas frente a estas materias. Es interesante advertir que 24 instituciones (46,15%) declaró que sus repositorios fueron implementados para datos de investigación antes de 2010, arrojando una antigüedad superior a los 5 años. Por el contrario, otras 25 (48,08%) reconocieron que éstos son relativamente nuevos, siendo implementados a partir de 2016 a la fecha. En tanto, apenas 3 instituciones (5,77%) afirmaron que sus repositorios fueron desarrollados en el período comprendido entre 2010 y 2015. En otras palabras, esta curva nos estaría demostrando que hubo un auge de repositorios durante la década

pasada, luego una desaceleración durante la primera mitad de esta década, repuntando este tipo de tecnología nuevamente en estos últimos años.

Gráfico 9-15: Antigüedad de repositorios

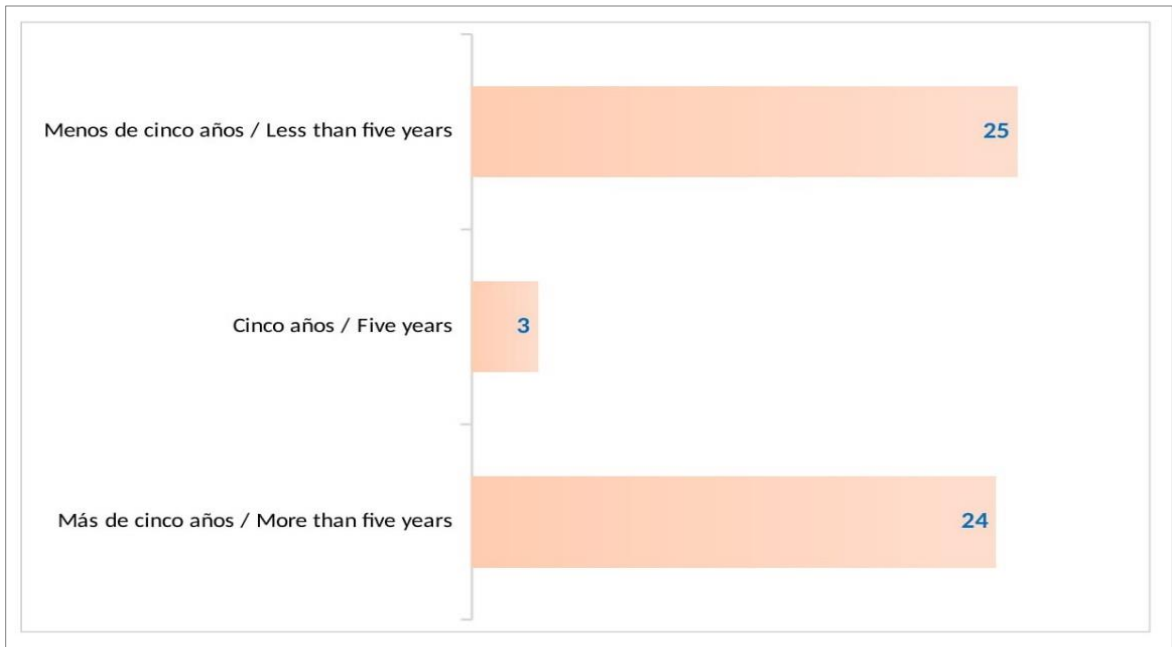


Gráfico 9-16: Antigüedad de repositorios

6. ¿Hace cuánto tiempo el repositorio de su org ...		
Respuesta	Número de envíos	% de envíos totales
Menos de cinco años / Less than five years	25	48,08%
Cinco años / Five years	3	5,77%
Más de cinco años / Más de cinco años	24	46,15%
Total	52	100,00%

Respuestas totales: 52

Pregunta 7:
¿Con qué tipo de financiamiento cuenta el repositorio y/o la sección de datos de investigación?

Sin lugar a dudas, el aporte de las organizaciones sin fines de lucro ha sido esencial en el desarrollo de los repositorios de datos, como se demuestra -además- por el hecho que la mayor parte de éstos se encuentra en los países llamados “del primer mundo”, entre los cuales se cuentan varios de Europa, EE.UU., Canadá y Australia. En estos países, además, hay políticas gubernamentales que estimulan su creación, al tiempo que ofrecen un marco jurídico regulatorio que promueve no sólo la colaboración, sino también las buenas prácticas. Lo anterior se sustenta en que, de las 52 instituciones encuestadas, 32 (61,54%) respondieron que el financiamiento proviene de sus propios presupuestos administrativos. En segundo lugar, apenas 7 (13,46%) aseguraron no tener financiamiento, mientras que sólo 6 (11,54%) reconocieron que éste provenía de las arcas fiscales.

En los países menos desarrollados, ésta podría ser una causa que explicaría la lenta o incipiente implementación de los repositorios de datos, ya que los gobiernos suelen poner los limitados recursos económicos disponibles en el desarrollo de políticas sociales por sobre políticas que favorezcan a la academia y/o a la investigación. Del mismo modo, la opción de obtener fondos por medio de la adjudicación de proyectos pareciera ser la menos viable para la consecución de este objetivo, con apenas 5 instituciones (9,62%) que reconocieron esta modalidad de financiamiento, porque en palabras de los investigadores toma mucho tiempo, es costoso porque tiene que encontrar el equilibrio con otros proyectos y es poco probable que sea sustentable en el tiempo.

Gráfico 9-17: Tipo de financiamiento

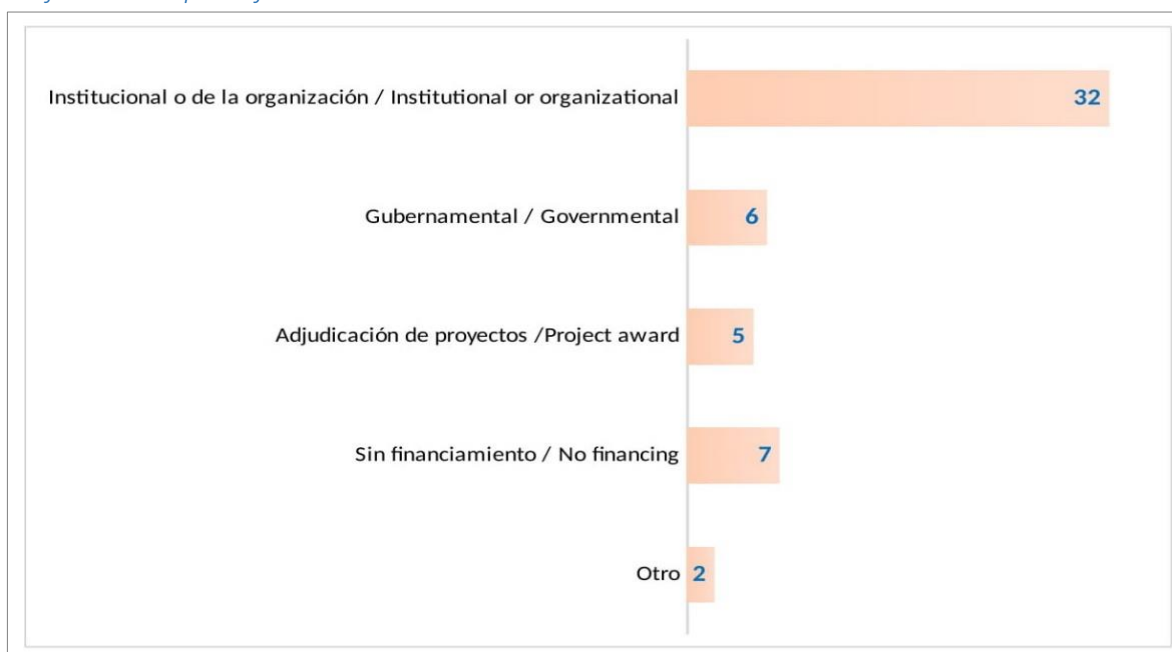


Gráfico 9-18: Tipo de financiamiento

7. ¿Con qué tipo de financiamiento cuenta el r ...

Respuesta	Número de envíos	% de envíos totales
Institucional o de la organización / Institucional u organizacional	32	61,54%
Gubernamental / Gubernamental	6	11,54%
Adjudicación de proyectos / Premio proyecto	5	9,62%
Sin financiamiento / Sin financiamiento	7	13,46%
Otro	2	3,85%
Total	52	100,00%

Respuestas totales: 52

Pregunta 8:

¿Cuán satisfechos están actualmente con el software del repositorio?

Las respuestas de las instituciones que participaron en la encuesta evidencian que, en general, la experiencia del software utilizado en los repositorios ha sido al menos buena, evaluación que da cuenta de aspectos como la gestión del mismo, claridad para acceder a la información, amigabilidad de la página (tanto para los administradores como para los usuarios), capacidad de almacenamiento e interoperabilidad. Aquí también se manifiesta la presencia de los principios FAIR en este punto. De esta forma, es elocuente que 25 instituciones (48,08%) se hayan manifestado **Muy satisfechas**; seguidas de 13 (25%) que se encuentran **Algo satisfechas** y de 9 (17,31%) que están **Extremadamente satisfechas**. Así, poco más del 90% del total de quienes participaron en la encuesta reportó una experiencia en general positiva y estimulante, evaluación que dice relación con lo necesarias que se han vuelto estas tecnologías. Reforzando esta idea, resulta muy llamativo que apenas 4 instituciones (7,69%) declararon estar **No tan satisfechas** con el software, mientras que sólo 1 (1,92%) calificó su experiencia como **Nada satisfechas**. Con todo, estas últimas representan el 9,61% del total, evidenciando que los softwares para repositorios de datos disponibles en la actualidad se han diseñado de manera funcional para servir a las necesidades de los usuarios, lo que debería estimular la capacitación del personal de las instituciones y predisponerlos favorablemente ante dicha experiencia.

Gráfico 9-19: Satisfacción con el software

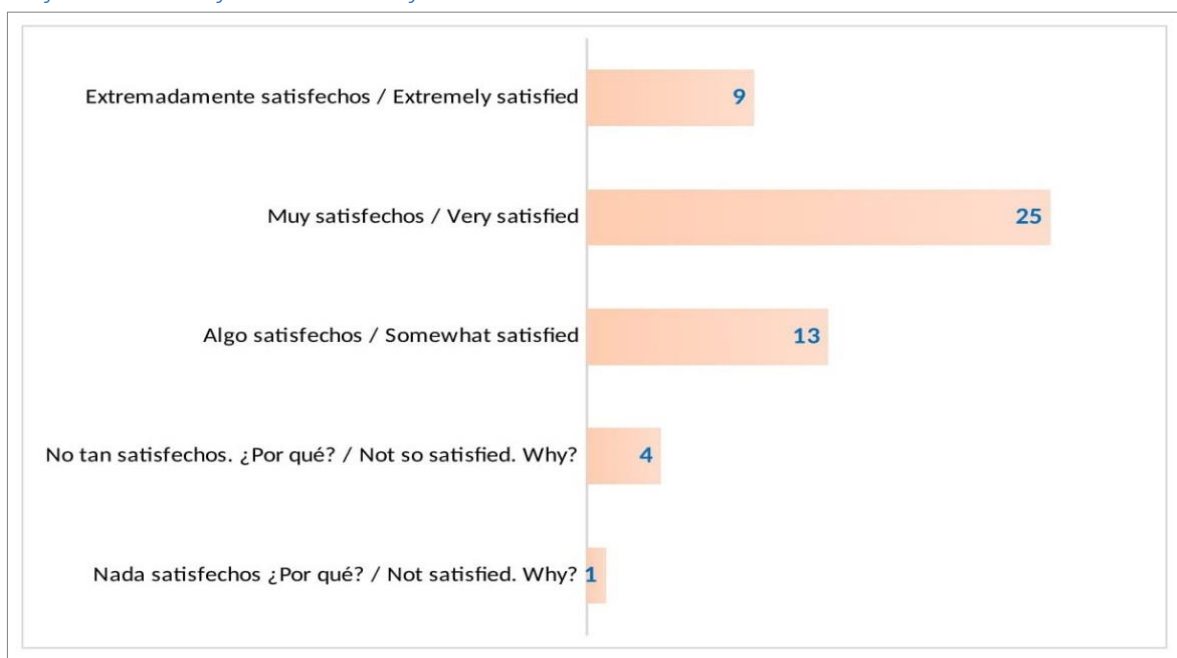


Gráfico 9-20: Satisfacción con el software

Respuesta	Número de envíos	% de envíos totales
Extremadamente satisfechos / Extremadamente satisfecho	9	17,31%
Muy satisfechos / Muy satisfecho	25	48,08%
Algo satisfechos / Algo satisfecho	13	25,00%
No tan satisfechos. ¿Por qué? / No tan satisfecho. ¿Por qué?	4	7,69%
Nada satisfechos ¿Por qué? / No satisfecho. ¿Por qué?	1	1,92%
Total	52	100,00%

Respuestas totales: 52

Pregunta 9:

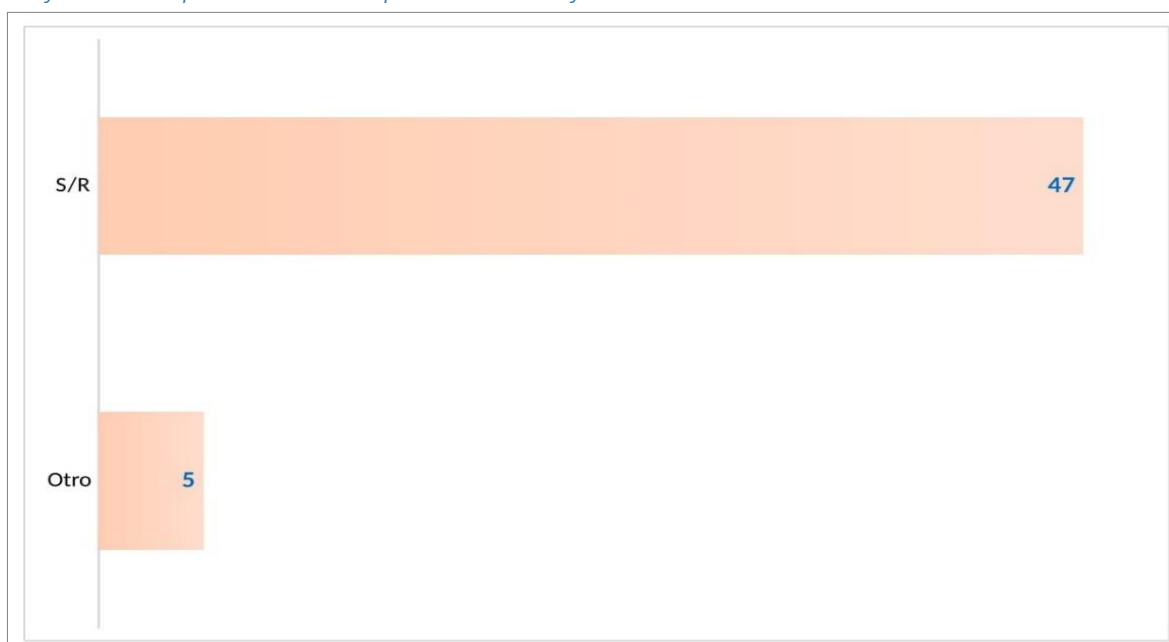
¿Por qué su respuesta en la pregunta anterior es "No tan satisfecho" o "Nada Satisfecho"?

Considerando que las respuestas a la pregunta anterior, con la cual ésta se relaciona, fueron en su mayoría buenas, 47 instituciones (90,38%) se abstuvieron de hacer comentarios sobre lo ya declarado, a excepción de 5 (9,62%), que entregaron sus argumentos para justificar su postura:

- Alta carga de mantenimiento del software.
- Fin de vida útil del repositorio sin soporte para los desarrolladores.
- Lentitud de la operabilidad del software y desactualización de la versión en uso.
- Software de código abierto, con soporte deficiente y actualizaciones futuras de difícil implementación.

Básicamente, estos comentarios dejaron en evidencia que el software es el responsable de la deficiente evaluación del repositorio de datos, de tal manera que -si se implementaran mejoras oportunas, económicas, de baja complejidad y de fácil gestión- las 5 instituciones descontentas podrían modificar su percepción respecto del software y su impacto en la administración de los repositorios de datos de investigación. Por otra parte, queda de manifiesto como buena práctica a considerar la relevancia que debe tener la elección de un software, al momento de implementar un proyecto de esta naturaleza. La sugerencia es buscar exhaustivamente en el mercado, evaluando más de una opción.

Gráfico 9-21: Opiniones sobre experiencia con software



Pregunta 10:

¿Cuán satisfechos están con la seguridad de los datos de investigación almacenados en el repositorio?

Al igual que en la Pregunta 8, en ésta las respuestas de las instituciones encuestadas evidenciaron una percepción favorable sobre la seguridad de los datos de investigación almacenados en el repositorio, llegando al 94,23%. Con 22 instituciones (42,31%), la opción **Muy satisfechos** se impuso con una distancia moderada respecto de la siguiente opción, **Extremadamente satisfechos**, que concentró 14 respuestas (26,92%), seguida de 13 (25%) que optaron por **Algo satisfechos**. Sólo 2 instituciones (3,85%) emitieron una evaluación más bien negativa respecto de este punto, **No tan satisfechos**, mientras que 1 (1,92%) no emitió respuesta y no hubo ninguna que se manifestara **Nada satisfechos**. De esta manera, puede afirmarse que la seguridad de los datos de investigación en los repositorios no es un tema que inquiete a las instituciones que respondieron, las cuales evidenciaron más bien lo contrario.

Gráfico 9-22: Seguridad de datos

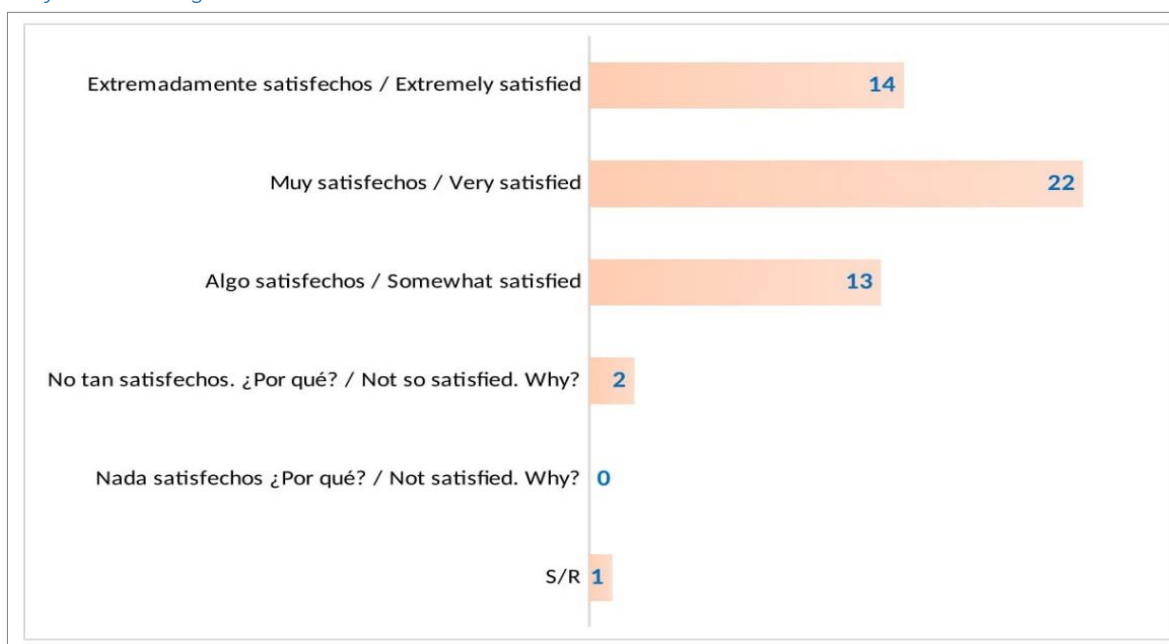


Gráfico 9-23: Seguridad de datos

10. ¿Cuán satisfechos están con la seguridad d ...

Respuesta	Número de envíos	% de envíos totales
Extremadamente satisfechos / Extremadamente satisfecho	14	27,45%
Muy satisfechos / Muy satisfecho	22	43,14%
Algo satisfechos / Algo satisfecho	13	25,49%
No tan satisfechos. ¿Por qué? / No tan satisfecho. ¿Por qué?	2	3,92%
Total	51	100,00%

Respuestas totales: 51

Pregunta 11:

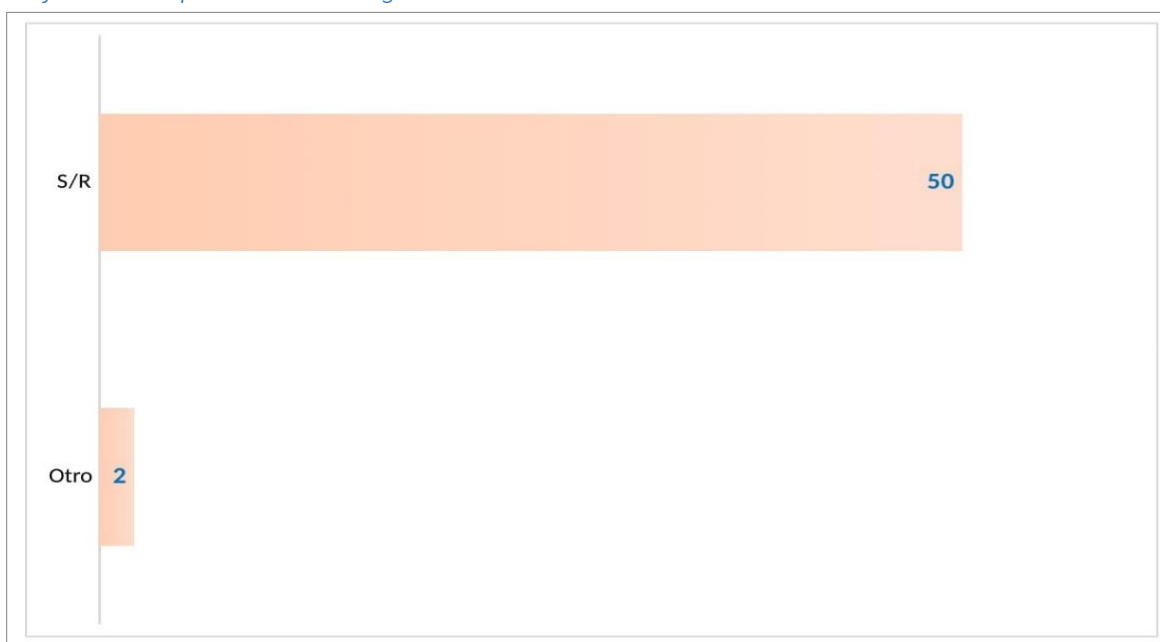
¿Por qué su respuesta en la pregunta anterior es "No tan satisfechos" o "Nada Satisfechos"?

Dado que sólo 2 instituciones declararon sentirse **No tan satisfechos** (3,85%), mientras que ninguna evaluó con **Nada satisfechos** y 1 (1,92%) se abstuvo de emitir una opinión, en esta pregunta sólo se entregaron los siguientes argumentos para validar las calificaciones emitidas anteriormente:

- Migración hacia un soporte más seguro, ya que la seguridad no se encuentra garantizada por el momento (ausencia de SSL²³⁶, gestión deficitaria de usuarios, entre otros).
- Diseño deficitario del repositorio para proteger los datos y/o su acceso a ellos.

²³⁶ Secure Sockets Layer (SSL) o capa de sockets seguros, correspondiente a una tecnología estándar para mantener segura una conexión a Internet, protegiendo cualquier información confidencial enviada entre dos sistemas con el fin de impedir que terceras personas modifiquen algún dato transferido. En esto último se considera, incluso, información de tipo personal. Más datos en: <https://www.websecurity.digicert.com/es/es/security-topics/what-is-ssl-tls-https>

Gráfico 9-24: Opiniones sobre seguridad de datos



Pregunta 12:

¿Cuán satisfechos están con la confiabilidad en relación a criterios de auditoría y certificación del repositorio de datos?

44 instituciones (84,62%) declararon sentirse confiadas respecto de este tema. Esta cifra se desglosa de la siguiente manera: 25 (48,08%) seleccionaron **Muy satisfechos**; 16 (30,77%), **Algo satisfechos**; y 3 (5,77%), **Extremadamente satisfechos**. Sólo 6 instituciones (11,54%) manifestaron una evaluación negativa, donde 4 (7,68%) se inclinaron por **No tan satisfechos**, mientras que 2 (3,85%) por **Nada satisfechos**. En tanto, 2 instituciones (3,85%) no emitieron opinión frente a este requerimiento. En líneas generales, estas respuestas demostraron, nuevamente, una percepción positiva en otro de los aspectos relacionados con los repositorios de datos, como es la confiabilidad en cuanto a criterios de auditoría y certificación. Lo anterior habla no sólo del alto grado de gestión alcanzado por los desarrolladores, sino también de la constante fiscalización o curatoría ejecutada a través de diversas instancias para garantizar el correcto funcionamiento y las buenas prácticas dentro de las comunidades que emplean los repositorios de datos de investigación. Queda de manifiesto, así, una permanente capacidad de mejora por parte de los desarrolladores, al igual que protocolos más profesionales y complejos, generados por las instituciones o las autoridades, para asegurar un marco de funcionamiento más confiable y seguro.

Gráfico 9-25: Satisfacción con confiabilidad

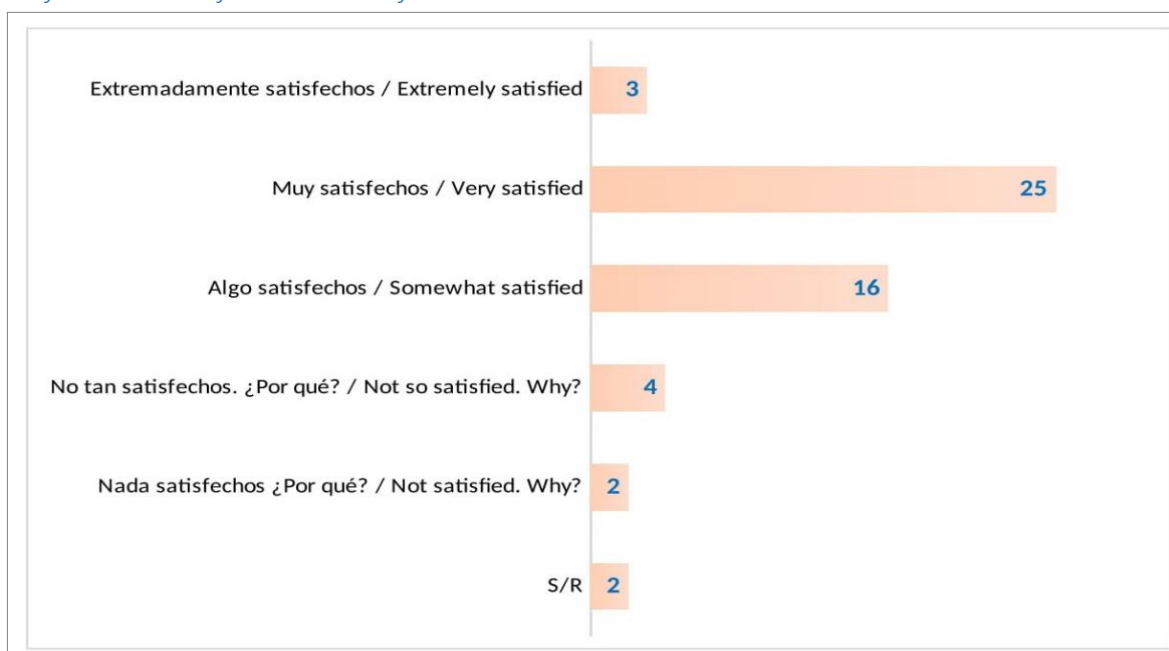


Gráfico 9-26: Satisfacción con confiabilidad

Respuesta	Número de envíos	% de envíos totales
Extremadamente satisfechos / Extremadamente satisfecho	3	6,00%
Muy satisfechos / Muy satisfecho	25	50,00%
Algo satisfechos / Algo satisfecho	dieciséis	32,00%
No tan satisfechos. ¿Por qué? / No tan satisfecho. ¿Por qué?	4	8,00%
Nada satisfechos. ¿Por qué? / No satisfecho. ¿Por qué?	2	4,00%
Total	50	100,00%

Respuestas totales: 50

Pregunta 13:

¿Por qué su respuesta en la pregunta anterior es "No tan satisfecho" o "Nada Satisfecho"?

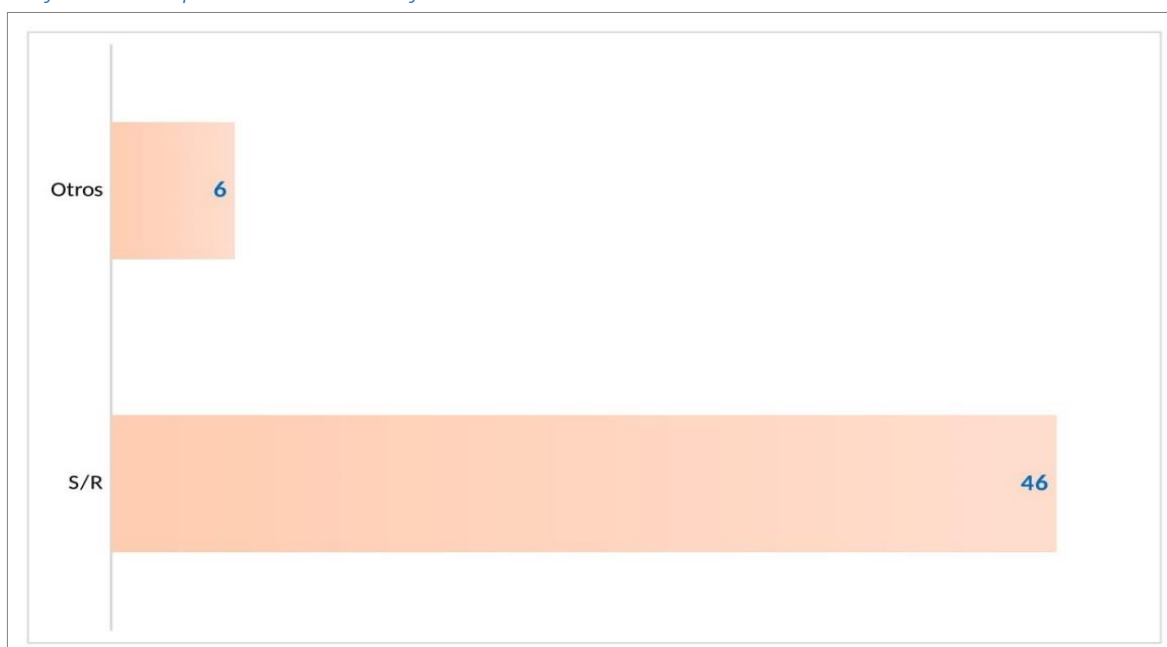
Considerando que en la pregunta anterior sólo 6 instituciones (11,54%) manifestaron una evaluación negativa (4 optaron por **No tan satisfechos** y 2 por **Nada satisfechos**, mientras que sólo 2 no emitieron opinión frente a este requerimiento), en este análisis se comparten los argumentos sobre los cuales se sustentaron dichas evaluaciones:

- Falta de iniciativa para buscar certificación, aunque las auditorías realizadas internamente han permitido detectar errores de corrección y algunos problemas de la capa de almacenamiento de preservación, los cuales podrían afectar la confiabilidad.
- Imposibilidad de obtener la acreditación OpenAire.

- Auditoría o certificación pendientes. “Sabemos que hay aspectos de nuestra implementación que no resistirían especialmente bien en un proceso de este tipo, aunque nos gustaría remediarlo”.
- Ausencia de criterios de auditoría y certificación.
- Auditoría o certificación pendientes. “Estamos interesados en el CoreTrustSeal²³⁷, pero no la hemos realizado por el coste que conlleva”.
- “No hemos realizado ningún proceso de certificación”.

De estas seis últimas preguntas, se destaca en consecuencia que las instituciones entregaron una alta ponderación a la confiabilidad del software (relacionada con criterios de auditoría y de certificación), como también a la seguridad y al software propiamente tal (con 28 instituciones de 52, 36 de 52 y 34 de 52, respectivamente, considerando las opciones **Extremadamente satisfechos** y **Muy satisfechos**). Por consiguiente, son aspectos que se sugiere sean considerados por parte de las instituciones que tengan planeado implementar o mejorar el repositorio de datos ya existente.

Gráfico 9-27: Opiniones sobre satisfacción



Pregunta 14:

Indique los aspectos más complejos que tuvieron que abordar durante la implementación del software o sección para datos de investigación

²³⁷ Es una herramienta de certificación de nivel básico basada en el catálogo y en los procedimientos de DSA-WDS Core Trustworthy Data Repositories Requirements que garantiza confiabilidad a los repositorios. Sitio web: <https://www.coretrustseal.org/about/>

En cuanto a esta pregunta, 39 instituciones (75%) comentaron los aspectos más complejos de la implementación del software, que se reproducen en la tabla incluida a continuación del gráfico, mientras que 13 (25%) se abstuvieron de formular contenidos válidos.

Gráfico 9-28: Aspectos complejos en implementación de software



A continuación, se presentan los 39 comentarios (75%) formulados en relación a esta pregunta de la encuesta, señalando que las respuestas en inglés fueron traducidas al español para uniformar el criterio idiomático, y reproduciéndolos textualmente.

Tabla 9-1: Aspectos complejos en implementación de software

	COMENTARIOS:
1	Gestión de Embargos. Integración con otros sistemas como ORCID. Estructura del formulario de envío. En desarrollo aún la implementación y difusión de servicio.
2	Los inconvenientes fueron poder ponernos de acuerdo con el personal de cómputo para poder instalar y desplegar los datos, así como el tipo de datos, su estructura y metodología que conlleva cada uno de ellos, que fue en el Área de Ciencias de la Tierra.
3	DSpace no es un software fácil de instalar.
4	Al inicio, como no se conocía mucho sobre repositorios, fue algo nuevo el ir conociendo la herramienta, la implementación con el software que se necesita para su funcionamiento era desconocido, pero sobre la marcha se fue afianzando y reforzando la experiencia del uso del repositorio.

5	Al parecer no hubo problemas, es decir, el desarrollador no reportó incidentes al respecto.
6	Integración de distintos estándares.
7	Fue una implementación bastante simple. Nada especial que reseñar.
8	La adaptación de los datos con respecto a lo solicitado por las autoridades correspondientes.
9	Las modificaciones al DSpace necesarias para la interoperabilidad con el repositorio nacional de CONACYT.
10	No tenemos una sección para datos de investigación.
11	La recuperación de contenidos de la institución.
12	La curva de aprendizaje.
13	La organización fue quien afrontó todos los aspectos de la gestión.
14	Tratar con conjuntos de datos mal implementados es un desafío. Además, los usuarios o el remitente a menudo no quieren hacerlo o se sienten algo obligados a hacerlo, por lo que no ponen mucho esfuerzo en la forma en que envían sus datos, lo que refleja la calidad de los conjuntos de datos (la mayoría de los campos opcionales están vacíos, el mínimo estricto en los metadatos y campos obligatorios). Además, tratar con conjuntos de datos heterogéneos (diferentes campos, diferentes trabajos) también es un desafío, ya que algunos conjuntos de datos son bastante complejos y, en algún momento, nuestro software tal vez no es lo suficientemente flexible para aprovecharlo al máximo.
15	Se debe realizar una planificación exhaustiva antes de la ejecución.
16	Respaldar flujos de trabajo para disertaciones / tesis, respaldar el acceso basado en roles de usuario, respaldar a coautores de terceros que no están afiliados a nuestra universidad, con nuestros objetos digitales como artículos, datos, código, desarrollo del lado de escritura de nuestra API ²³⁸ .
17	Creo que los aspectos más complejos tienen que ver con los procesos de almacenamiento y preservación, para evitar que los formatos pierdan vigencia, junto con especificar correctamente la estructura de metadatos.
18	La creación de herramientas de curador para la ingesta y la edición ha sido quizás uno de los mayores desafíos.
19	Dado que los materiales que se reúnen son documentos fuentes de diversa naturaleza (texto, audio, vídeo), fue necesario investigar cómo tratarlos previamente y cómo hacer que el software los visualice adecuadamente.
20	El tipo de datos con los que tratamos cambia todo el tiempo, por lo que el software debe actualizarse a intervalos regulares. Esto se hace más fácil al ser desarrollado internamente.
21	El esquema de metadatos implementado fue creado por el financiador del proyecto y es muy complejo y no se usa fuera de Australia. Esto ha significado que hemos tenido

²³⁸ Application Programming Interface (API) es una interfaz de programación de aplicaciones o conjunto de rutinas que provee el acceso a funciones de un determinado software y son publicadas por los constructores de éste. Son utilizadas para construir sus aplicaciones sin necesidad de volver a programar funciones ya realizadas por otros, reutilizando códigos válidos y operativos. Sitio web: <https://maplink.global/es/blog/como-usar-apis/>

	que simplificar el esquema para nuestros usuarios, con el fin de asegurarnos de que el proceso de depósito sea mucho más fluido. VIVO ²³⁹ es un software de código abierto para perfiles de personal y realmente no era adecuado para su propósito, por lo que para superar esto, tuvimos que hacer mucha personalización.
22	El conocimiento en sí de los datos de investigación como material. Desde la biblioteca, gestores del repositorio, conocemos mucho más a fondo las publicaciones.
23	Nuestra versión de DSpace (la 5.5) permite añadir vocabularios, así nuestros informáticos tuvieron que hacer poco trabajo de ajustes. Con los nuevos vocabularios (DataCite y COAR) estamos haciendo tests de recolección para la optimización de la semántica y la máxima fairness desde los criterios de Driver 2.0. Tenemos planificado poder exponer vía OAI-PMH con Driver 4.0 y así cumplir con las buenas normas de OpenAire 4.
24	Ser capaz de personalizar el software y, aun así, aprovechar el aspecto de soporte de la comunidad y los nuevos lanzamientos de versiones.
25	Mantener un soporte constante y de alta calidad para nuestro software y base de datos, especialmente diseñados, ha sido un desafío.
26	Adaptación del software para la automatización de publicaciones de grandes conjuntos de datos. Mantenimiento general de un sistema antiguo.
27	El aspecto más complejo fue la integración con el almacenamiento.
28	Alinear todo el software para que funcione como se espera y de forma segura.
29	Los aspectos más complejos fueron migrar datos de la plataforma de repositorio anterior, conectar la plataforma de almacenamiento (basada en la nube) al almacenamiento local y comunicar la existencia del nuevo repositorio a la comunidad de investigación institucional.
30	La parte más compleja de la implementación sigue siendo convencer a profesores e investigadores sobre la importancia de compartir sus datos de investigación, así como demostrar las ventajas y beneficios de la reutilización de datos.
31	Desarrollar un modelo de servicio y/o soporte sostenible con recursos mínimos.
32	La mayor complejidad está en el sistema de exploración de datos, no en el archivo.
33	(Fórmula) = - Uso de metadatos especializados - Gestión de identificadores DOI para conjuntos de datos.
34	Nuestro repositorio de datos geoespaciales está construido a partir de varios componentes independientes de código abierto. Necesitábamos descubrir cómo hacer que funcionen juntos como un sistema integrado. La otra parte compleja era convertir nuestros datos y metadatos del sistema anterior de 20 años.
35	Adecuación de formularios, metadatos, licencias y exportación a la tipología dataset.

²³⁹ VIVO es un software de código abierto que admite la grabación, edición, búsqueda, navegación y visualización de la actividad académica, fomentando la exhibición de este tipo de registros, el descubrimiento de la investigación y la evaluación de su impacto, entre otros. Sitio web: <https://duraspaces.org/vivo/about/>

36	Aprender toda una nueva filosofía, al ser el primer proyecto que trabajaba sobre una plataforma Big data (Hadoop) ²⁴⁰ . El número de componentes a gestionar es grande y la complejidad de todo el sistema también.
37	El repositorio de datos se utiliza sólo para datos anonimizados. No cumple con los requisitos de GDPR ²⁴¹ para el almacenamiento de datos confidenciales. No hemos intentado obtener una certificación externa para el repositorio.
38	La compatibilidad con las directrices de OpenAire en un principio, pero ya las ofrece por defecto el software en las nuevas versiones.
39	Usamos el mismo repositorio que para las publicaciones. Esta solución carece de algunas opciones de otros repositorios dedicados únicamente a datos, pero resultó más sencilla de implementar.

Visto lo anterior, estas respuestas pueden ser de gran utilidad para instituciones que estén planeando implementar o mejorar la gestión de un repositorio de datos de investigación. En este contexto, se podrían agrupar los comentarios proporcionados, utilizando alguna técnica o herramienta para estos fines. Sin embargo, a modo de conclusión, se puede señalar en términos generales que los aspectos más complejos se inclinan a la elección del **software** -un elemento estratégico-, el cual lidera las menciones en la tabla, convirtiéndose en un actor principal para implementar un proyecto de esta naturaleza. Su elección es de gran relevancia y deben considerarse instancias desde un buen soporte hasta la complejidad o cantidad de migraciones a nuevas versiones que requerirá, así como también potenciales dificultades que puedan surgir al momento de realizar integraciones con identificadores persistentes. El **desarrollo** propiamente de la tecnología y de la plataforma también son aludidos. En dicho sentido, llama la atención una expresión referida a “la mayor complejidad en el sistema está en la exploración de los datos, no en el archivo”, lo que se vincula a otras menciones relativas a la **recuperación**. Asimismo, un par de comentarios surgen en relación a la “**curva de aprendizaje**” que requiere una implementación de este tipo. Finalmente, se destaca que se manifestaron aspectos concernientes a la **planificación** necesaria, al **respaldo de flujos** y todas las tareas de gestión, actividades previas, como garantes de un buen desarrollo tecnológico. En este caso, como se mencionó, estas fases fueron identificados como elementos complejos al momento de responder a la encuesta y deberían ser registrados para considerar.

²⁴⁰ Hadoop es una estructura de software de código abierto para almacenar datos y ejecutar aplicaciones en clústeres de hardware comercial, proporcionando almacenamiento masivo para cualquier tipo de datos, y que se caracteriza por tener un enorme poder de procesamiento, como también la capacidad de procesar tareas o trabajos virtualmente ilimitados. Sitio web: https://www.sas.com/es_cl/insights/big-data/hadoop.html

²⁴¹ Reglamento General de Protección de Datos (o General Data Protection Regulation, GDPR), por medio del cual el Parlamento Europeo, el Consejo de la UE y la Comisión Europea tienen la intención de reforzar y unificar la protección de datos para todas las personas dentro de la UE. Su principal objetivo es dar control a los ciudadanos y residentes sobre sus datos personales y simplificar el entorno regulador de los negocios internacionales, unificando la regulación dentro de la UE. Sitio web: <https://www.powerdata.es/gdpr-proteccion-datos>

Pregunta 15:

Comparta brevemente las principales recomendaciones (o buenas prácticas) para las bibliotecas que están implementando un repositorio o una sección de datos para investigación.

En esta pregunta, se midió una alta tasa de participación por parte de las instituciones, ya que 44 (84,62%) compartió sus experiencias en la implementación y sostenimiento de sus propios repositorios de datos, por medio de recomendaciones o buenas prácticas para orientar a otras instituciones que se encuentran aún en un estado de desarrollo incipiente con respecto a este tema. Es interesante ver cómo hay aspectos determinados por la envergadura del repositorio de datos propiamente tal, o por el software adquirido para este propósito, así como también la importancia que se debe dar a la capacitación del personal encargado de administrar estos repositorios. Asimismo, se describe la renuencia o falta de práctica por parte de los investigadores no sólo en cuanto a internalizar la práctica de compartir sus investigaciones, sino también los procedimientos administrativos que esta labor conlleva, como es completar los datos de rigor para realizar un registro pormenorizado de los datos. A continuación, la tabla reproduce textualmente todos los comentarios (se han traducido al español aquellos cuya versión original es en inglés).

Tabla 9-2: Recomendaciones y buenas prácticas

	COMENTARIOS:
1	Conocer la realidad de los datos que se están trabajando en las diversas áreas de la Universidad, dependiendo del tipo de dato se requiere un determinado metadato que permita describirlo adecuadamente y gestionarlo eficientemente. Que el repositorio de datos cuente con la capacidad de gestionar los datos según los principios FAIR. Promover el repositorio y el servicio entre investigadores(as) y académicos(as). Capacitar a aquellos bibliotecarios(as) que estarán a cargo de la curaduría de los metadatos y del asesoramiento a los académicos(as). Puede ser a través del curso de gestión de datos de investigación que se dicta en el MPGI o a través de cursos de Coursera o Elsevier.
2	Tener un equipo interdisciplinario: bibliotecarios, investigadores, informáticos.
3	a) Citación y procedencia en publicaciones académicas "cuando se considere necesario". Cuando los investigadores utilizan datos creados por otros, éstos deben ser citados con referencia a su autor, a su procedencia y a un identificador digital permanente; b) Interoperabilidad en la mayor medida posible. Tanto los datos de investigación como los metadatos que permiten la evaluación y reutilización, todo ello con el objetivo de que el intercambio y el enlace de plataformas e información permita un flujo transparente para el usuario; c) Reutilización no restrictiva. Los datos para investigación que están en el dominio público deben ser etiquetados como reutilizables por medio de una renuncia a los derechos o una licencia no restrictiva mediante la cual se deje en claro que los datos pueden ser reutilizados sin mayor requisito que el reconocimiento al autor; d) La información debe llevar el dataset o el dato del lugar donde se hospedarán "que se incluirá

	en el Repositorio dentro del archivo llamado Readme File ²⁴² con la siguiente información y descripción": i) Proporcionar la información sobre los datos y/o el set de datos para que sea correctamente interpretado tanto por personas como máquinas y se puedan realizar procesos con otras investigaciones. ii) Cuando el registro (unidad de descripción) se compone de varios sets de datos, se debe vincular a un archivo que se llame Readme File (archivo que contendrá la descripción de los datos), de tal manera que pueda ser asociado al set de datos y se especifique en el archivo de formato de fácil consulta como: csv o txt. iii) Debe incluir una breve descripción de los datos según sea el caso: contacto del investigador principal, fecha de recolecta de datos y de creación del set de datos.
4	Contar con un equipo que incluya tanto bibliotecólogos como informáticos.
5	Hacer el estudio previo de lo que se necesita para su repositorio, tanto de lo que se va a publicar como el software que van a elegir para su implementación, el solo hecho que quieren un repositorio e instalarlo sin tener conocimiento del alcance, a veces acarrea más problemas que beneficios, conocer las reglas que se necesitan para la implementación son de gran ayuda a la hora de ir definiendo las comunidades y colecciones, para no revolver y perderse en lo que se quiere hacer.
6	Llevar a cabo una buena planeación, que gire en torno las cualidades de interoperabilidad, escalabilidad y granularidad.
7	Utilice estándares de (meta) datos ampliamente adoptados, garantice la interoperabilidad con otros sistemas, porque las tecnologías cambian con el tiempo, por lo que cualquier repositorio/software siempre necesitará mantenimiento y actualizaciones (esto debe considerarse en la financiación), trabaje con los sistemas existentes cuando sea posible, en lugar de construir desde cero.
8	Personalmente, prefiero instalar un repositorio de software libre y pagar salarios de desarrolladores para mantenerlo y actualizarlo. Esto da la posibilidad de implementar mejoras funcionales en participación con la comunidad (importante siempre coordinarlo con la comunidad ya que cambios en el software que no se incluyan en la distribución oficial pueden llevar a incompatibilidades con versiones futuras). En cualquier caso, hay que disponer de técnicos informáticos debidamente cualificados.
9	Establecer un mecanismo para crear una cultura de compartir la información. Crear una política de cesión de derechos de autor. Revisar los requerimientos de la biblioteca, caso del uso de software libre, analizar si éste se adapta con facilidad a las necesidades. Garantizar el mantenimiento del software. Establecer políticas de seguridad que garanticen la integridad de los datos, así- como el acceso al servidor.
10	No pagar a proveedores, que lo hagan ellos.
11	Nuestro Repositorio Institución tiene en sus colecciones producción científica, académica y patrimonial.

²⁴² Los archivos Readme suelen crearse como readme.txt o readme.md y suelen contener información importante sobre el sistema, proyecto o software al que se refieren. Para que los usuarios puedan encontrar fácilmente el archivo de un vistazo, se recomienda ubicarlo en el nivel superior del directorio. Tienen diferentes propósitos según los distintos tipos de usuarios. Sitio web: <https://www.ionos.es/digitalguide/paginas-web/desarrollo-web/archivo-readme/>

12	El proyecto debe ser apoyado por la alta dirección de la institución y organización, se debe contar con una definición de contenidos a almacenar identificando previamente lo que tiene disponible la institución, contar con un equipo técnico para el uso y administración del tema tecnológico, entre otros.
13	Tomar el tiempo para seleccionar el software, que este se adecue a las necesidades de la institución. La mejor recomendación es implementar uno de open source; además de libre disponibilidad, su código fuente permite "corregir" errores, modificarlo e integrar desarrollo propio. En el equipo de trabajo incluir personal capacitado en las dos vertientes: administrador del repositorio, y (sic).
14	Estudien bien los casos existentes y sus propias necesidades, manténganse atentos a la interoperabilidad entre sistemas y sobre todo a nuevas funcionalidades que atiendan nuevas necesidades.
15	Esto debe ser sencillo para un usuario que puede no estar motivado para hacerlo. Menús desplegables, palabras clave sugeridas, mucha comida con cuchara en el camino. También me gustaría preparar una plantilla por tema de investigación para ayudarlos a comenzar (para temas populares). Hacer vídeos (tutoriales de Youtube) parece ayudar también.
16	Es muy importante implicar a los investigadores para que autoarchiven sus datos en el repositorio.
17	Debe haber un monitoreo de calidad de los datos en el repositorio.
18	Ver: CORE TRUST SEAL https://www.coretrustseal.org/why-certification/requirements/ ; Consulte: Marco comunitario de buenas prácticas de COAR https://www.coar-repositories.org/coar-community-framework-for-good-practices-in-repositories/ (próximamente la versión en español); Ver: TRAC - Lista de verificación de archivo del repositorio confiable (TRAC) contiene métricas que ayudan a evaluar el repositorio en las áreas de administración y política, manejo de objetos y tecnología. (https://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/trac); Consulte también la norma ISO16363 de febrero de 2012 (http://www.iso.org/iso/catalogue_detail.htm?csnumber=56510) o TDR (http://public.ccsds.org/pubs/652x0m1.pdf) . Es una revisión de la lista de verificación TRAC. TDR es una versión gratuita y no oficial publicada por el Comité Consultivo de Sistemas de Datos Espaciales (CCSDS); y Ver también: Chodacki, John, Cynthia Hudson-Vitale, Natalie Meyers, Jennifer Muilenburg, Maria Praetzellis, Kacy Redd, Judy Ruttenberg, Katie Steen, Joel Cutcher-Gershenfeld y Maria Gould. Implementación de prácticas de datos efectivas: recomendaciones de las partes interesadas para el apoyo a la investigación colaborativa. Washington, DC: Association of Research Libraries, septiembre de 2020. https://doi.org/10.29242/report.effective-datapractices2020
19	Trabajar con esquemas de metadatos estándar para fomentar la interoperabilidad.
20	Comunicarse regularmente con los investigadores, si es posible, para mantenerse al tanto de sus necesidades cambiantes.

21	Definir los formatos de archivo que se aceptarán. Formar al profesional de la información en estos temas para que, luego, pueda asesorar a los grupos de investigación. Falta el hábito de especificar en las planillas de cálculo a qué corresponde cada columna/fila de datos volcados. Definir si el repositorio sólo proporcionará los conjuntos de datos (descargables) o también herramientas para su visualización. Relacionar los conjuntos de datos con los documentos en los que éstos fueron utilizados.
22	¡Utilice vocabularios y ontologías controlados para organizar los datos!
23	Utilice un esquema de metadatos ampliamente utilizado y evite las personalizaciones si es posible. Compre una suscripción a un servicio alojado, como Figshare ²⁴³ , en lugar de desarrollar su propio repositorio institucional.
24	Soporte a medida de las necesidades de cada investigador.
25	Utilizo de DATACITE como estándar de exposición de los metadatos, con particular atención a los Identificadores que se conectan al dataset, como handle o doi de artículos relacionados, o relación con el proyecto financiador.
26	No somos una biblioteca sino un archivo de datos.
27	Hacer uso de los procesos de auditoría existentes (por ejemplo, WDS) y organizaciones profesionales, para garantizar que se sigan las mejores prácticas.
28	A menos que tenga una gran experiencia técnica y soporte, necesita tener una muy buena relación con el área de su organización que está brindando soporte. Muchas instituciones piensan que los problemas que enfrentan son únicos y especiales y, por lo tanto, necesitan un sistema a medida. Estos sistemas son más difíciles de mantener, así que observe qué tan únicas son realmente sus necesidades (pista: probablemente no lo sean). El hecho de que nadie más en su institución esté almacenando datos no significa, necesariamente, que la biblioteca sea el grupo correcto para hacerlo. Pregúntese: “¿Es esto tan importante que debería hacerse de una manera posiblemente mediocre?”
29	Aunque aún no son visibles, estamos trabajando con conjuntos de datos no tradicionales, conformados por archivos de audio, planos, dibujos, documentos fonéticos, etc. para áreas como lingüística, arquitectura y humanidades.
30	Soporte para los principios de datos FAIR. Comunicación abierta y regular con socios de tecnología de la información.
31	No hemos tenido muchas interacciones con los repositorios de datos de la biblioteca, como un repositorio (marino) específico de dominio; sólo podemos sugerir que el repositorio trate de seguir los principios FAIR.
32	Elija software que tenga barreras bajas para la entrada y el uso por parte del investigador. Fomente el envío automático de contenido. Comunicar amplia y repetidamente la existencia del nuevo repositorio. Obtenga la aceptación de las partes interesadas institucionales, por ejemplo, la oficina de investigación, el liderazgo de la Facultad, los servicios centrales de TI para respaldar y promover el repositorio de datos.

²⁴³ Figshare es una plataforma utilizada para preservar y compartir los resultados las investigaciones realizadas por los investigadores, considerando figuras, conjuntos de datos, imágenes y vídeos. Fue lanzada en enero de 2011 por Mark Hahnel y cuenta con el apoyo de Digital Science, que hizo un relanzamiento en enero de 2012. Sitio web: <https://figshare.com/>

33	La biblioteca y los bibliotecarios deben tomar la iniciativa en la gestión de repositorios de datos, precisamente porque ya trabajan con datos en diferentes plataformas y formatos. Sería interesante orientar a los sistemas bibliotecarios universitarios para que inviertan en la formación de su personal, con el fin de adoptar el trabajo con datos para ayudar a los investigadores a gestionar, compartir y reutilizar los datos científicos. Lo ideal sería expandir los servicios de referencia que ya se ofrecen, ampliando las actividades para que sea un amplio centro de apoyo a la investigación, ofreciendo soporte en el descubrimiento de repositorios de datos, asistencia con temas de licencias de derechos de autor, preparación de datos, licenciamiento, conversión de datos, metadatos, depósito, métricas y ayuda con la reutilización.
34	Intente todo lo que pueda para convencer a las personas sobre la importancia y los beneficios de un repositorio de datos de investigación, ya que las personas generalmente se resisten a depositar sus datos.
35	Es más que un sistema. Se trata de cómo los investigadores lo van a utilizar. Tómese el tiempo para comprender eso y comprender los flujos de trabajo administrativos y de investigación asociados con los datos.
36	Capacidad de la solución para interactuar, directamente, con otros sistemas a través de API.
37	El sistema de archivos ZFS proporciona una suma de comprobación automática y una duplicación sencilla en el almacenamiento remoto. Escriba en cintas LTO para almacenamiento externo a largo plazo y agregará datos incrementales de escritura al almacenamiento de AWS Glacier. Utiliza un sistema de archivos estándar confiable en lugar de una base de datos compleja o interfaces sofisticadas, pero también proporciona una interfaz de búsqueda y lista de datos. La mayoría de los datos en formatos científicos de autodescripción (CDF, netCDF) con metadatos definidos, documentos y software clasificados en directorios junto con datos.
38	Fórmula) = - Implementar políticas de depósito - Implementar criterios de curación de contenidos - Analizar y acogerse a los principios FAIR.
39	Le recomiendo encarecidamente que aproveche la experiencia disponible a través de Research Data Alliance, en particular los grupos de Bibliotecas y Repositorios: https://www.rd-alliance.org/
40	Siga los estándares de metadatos. Encuentre una comunidad (como GeoBlacklight ²⁴⁴) de instituciones que estén haciendo cosas similares y trabajen juntos para lograr objetivos comunes.
41	Nuestro volumen actual de datasets es pequeño, por lo que lo más sencillo fue customizar DSpace para aceptar la tipología de datasets, no cuenta con todos los servicios que brinda, por ejemplo, Dataverse o un repositorio DSpace exclusivo para datasets, pero para el volumen actual cumple requisitos. Si el volumen crece, será necesario valorar otra opción.

²⁴⁴ GeoBlacklight es un software de colaboración multiinstitucional y de código abierto, que crea una mejor manera de buscar y compartir datos geospaciales. Sitio web: <https://geoblacklight.org/>

42	Para grandes volúmenes de datos (>1TiB), no lo duden, usen Hadoop. Usen formatos estándar (CSV, Parquet), y ofrezcan una manera fácil de descargar el dataset completo, sin tener que hacer una query del total.
43	Tenemos un grupo de trabajo que realiza el seguimiento del repositorio en el que hay bibliotecarios de todas las universidades del Consorcio Madroño, informáticos de la oficina técnica del Consorcio Madroño, un director de biblioteca y el director técnico del Consorcio Madroño. En este grupo de trabajo se han definido las políticas, metadatos mínimos, acuerdos sobre tipos de datos y quién puede subir datasets a los repositorios entre otros temas.
44	Trabajar conjuntamente con otros centros similares es básico, así como con los investigadores.

En esta tabla se manifiestan diversas recomendaciones y buenas prácticas que pueden ser de gran utilidad para instituciones que están planeando implementar o mejorar la gestión de un repositorio de datos de investigación. También se podrían agrupar los comentarios proporcionados, utilizando alguna técnica o herramienta para estos fines; no obstante, a modo de conclusión, los aspectos destacados dicen relación con que un proyecto de esta envergadura requiere de apoyo, posicionamiento, no sólo de instancias directivas, sino también de vinculación con investigadores a quienes se les debe comunicar, además, la existencia de esta tecnología, sus beneficios e implicancias. Una recomendación es, también, indagar estudios previos de lo que requerirá construir grupos multidisciplinarios. En dicho sentido, para conocer estudios de casos debidamente documentados se sugieren las experiencias de Research Data Alliance. Si bien en la tabla anterior se manifestaba como un problema complejo, acá surge nuevamente la elección del software a modo de buena práctica, al igual que la utilización de esquemas de metadatos que tiendan a la interoperabilidad, el uso de vocabularios controlados, el uso de estándares como DataCite, utilización de identificadores persistentes, adherir a los Principios FAIR y adoptar las buenas prácticas de COAR. Otro aspecto relevante es la sugerencia, en más de una oportunidad, de certificaciones para que su repositorio de datos sea una tecnología confiable, adoptando TRAC o Core Trust Seal y, por supuesto, disponer de políticas y criterios de curación de contenidos.

9.2. Resumen de buenas prácticas, recomendaciones y/o desafíos

En este apartado de capítulo, me refiero a las recomendaciones, buenas prácticas y/o desafíos relacionadas con la implementación y desarrollo de los repositorios de datos de investigación, basándome en la experiencia recogida a través de la metodología descrita anteriormente en esta publicación, sustentada sobre una exhaustiva revisión bibliográfica, una encuesta, entrevistas, la comunicación y colaboración digital con expertos en la materia, el análisis de dos casos de estudio y el benchmarking aplicado a algunas plataformas reconocidas internacionalmente en este tópico.

De este modo, aquí se incluye una tabla, clasificada por temáticas, tanto para buenas prácticas, como para recomendaciones, cuyo propósito es complementar y proporcionar una síntesis de la publicación.

Tabla 9-3: Principales buenas prácticas, recomendaciones y/o desafíos

Políticas – Principios - Licencias	
1	Sólidas políticas respaldadas en la legislación, luego de años de esfuerzos mancomunados para propiciar la Ciencia Abierta (CA) y el Acceso Abierto (AA) de carácter obligatorio, existen en los países de Europa al igual que en Estados Unidos y en otras potencias desarrolladas, las cuales se sustentan en que toda investigación financiada con recursos públicos debe ser compartida. En el resto del mundo -donde la CA y el AA aún se encuentran conquistando terreno- el avance es dispar, pero la visión está clara al menos y las legislaciones que se están promulgando van en esta misma dirección, como es el caso de Chile, donde la reciente consulta pública sobre estos aspectos espera fortalecer dicha temática. La recomendación -en este caso- es conocer el marco jurídico legal de su país y los elementos asociados a un proyecto de repositorios, incluyendo, pero sin limitarse, a los siguientes aspectos: a) Propiedad intelectual de los contenidos. b) Protección de datos personales de los usuarios del sistema. c) Relación entre privados y sector público, políticas anti corrupción y conflicto de intereses.
2	Respecto de licencias, una buena práctica es que los datos de investigación contenidos en los repositorios deben disponer siempre de opciones de adquisición de licencias y eventuales costos de soporte anual para obtener actualizaciones y soporte ante incidentes, lo menos restringidas en lo posible, CC-0 o CC-BY. Por ende, hay que conocer el alcance de este tópico para decidir el tipo de licencia más adecuada para el proyecto.
3	Los Principios FAIR propician que los datos sean ubicables, accesibles, interoperables y reutilizables, ya sea por el mismo investigador o por otros investigadores en cualquier parte del mundo. Estas características, que debieran ser inherentes a los datos de investigación (con independencia de su naturaleza, vale decir, si son datos duros, números, fotografías, infografías u otro contenido audiovisual), impactan favorablemente, incentivando la investigación y optimizando recursos económicos. Al momento de implementar o aplicar mejoras a un repositorio de datos, se recomienda conocer a cabalidad y adoptar estos Principios, los cuales -al materializarse- corresponden a buenas prácticas. Esto se verá impactado positivamente por el apoyo de profesionales con experiencia, como parte de los proyectos de implementación.
Gestión del Cambio - Cambio cultural	
1	El cambio cultural en temas de CA y AA involucra y sugiere aspectos tales como las políticas que adoptan los países, los acuerdos a los que llegan las organizaciones de investigación y el cambio que se debe generar en las prácticas de los investigadores. La invitación es a

	sumarnos, desde la tribuna que corresponda a cada cual, para aportar o incentivar paulatinamente este cambio. En el caso de los bibliotecarios, la recomendación y desafío es a prepararnos, adquiriendo conocimientos para compartir desde nuestro rol formativo, como un aporte para los investigadores, con el fin de mitigar -si existen- sus aprehensiones en favor de la postura, en muchos casos reticente aún, para educar en relación a temas de CA, AA y otras materias afines. Una buena práctica para nuestro quehacer será comprometernos con el aprendizaje y promoción de estos temas para informar, independientemente de las decisiones que se pueda adoptar, asumiendo con vehemencia que estamos frente a una oportunidad para vincularnos con el quehacer de investigadores y el ciclo de vida de los datos de investigación, entre otros aspectos. La gestión del cambio se debe considerar desde el inicio de cualquier proyecto, con roles claramente definidos de quienes asumirán dicho proceso, la estrategia para abordarlo y los recursos necesarios para su desarrollo. Este proceso de gestión del cambio no se debe subestimar y se debe asumir como un factor clave de éxito.
	Organizaciones - Instituciones
1	La recomendación es conocer el tipo de organización o institución en la cual se implementará un repositorio de datos de investigación. En este sentido, se recomienda conocer también la misión de la corporación donde éste se insertará, en el sentido de definir si se trata de una institución académica cuya misión es, por ejemplo, privilegiar la docencia por sobre la investigación. Si ese fuese el contexto, es probable entonces que no requiera un proyecto de esta naturaleza. En cambio, si su plan estratégico considera el incentivo de la investigación, entonces será indispensable que se implemente un repositorio de datos, o bien, si ya se cuenta con uno, introducir las mejoras indispensables de acuerdo con las directrices acordadas a nivel mundial, en la medida de lo posible. También se aconseja conocer la política o posición de la entidad frente a la CA y el AA. Esto también puede permitir, como buena práctica, definir el marco inicial de desarrollo y proyectar una adecuada carta de navegación. La cultura, madurez, misión y estrategia deben guiar la implementación de un repositorio. Su desconocimiento puede ser considerado un riesgo alto que impactará en el éxito del proyecto.
	Recursos Financieros
1	La recomendación en este caso es conocer y estar al día respecto de la comprensión que conlleva la CA y el AA, en el sentido no solamente de los beneficios sociales y culturales, sino también económicos. Los costos totales de implementación y de operación de un repositorio de datos de investigación deben ser evaluados y considerados en la aprobación del proyecto. Estos costos deben incluir software, consultoría, licencias adicionales, infraestructura computacional, mantenimiento de software y hardware, personal operativo, personal o contratos de soporte, entre otros.
2	Plan S de la Coalición S (integrada por 24 organizaciones con presencia mundial, lo cual claramente evidencia su carácter universal): una buena práctica para incentivar la producción de investigación científica y académica -lo cual también puede asumirse como un desafío- sería, por ejemplo, sumarse a un modelo parecido a éste e inyectar recursos públicos para incentivar la creación tanto de plataformas como de revistas de AA. Asociado a lo anterior, se

	postula que los costos de publicación de las investigaciones sean asumidos no por los autores (como sucede actualmente en muchos países del mundo), sino por las mismas universidades, institutos de investigación o patrocinadores, de acuerdo con un presupuesto estándar según las políticas internas e intereses estratégicos de las organizaciones.
3	A nivel de proyecto, una recomendación y buena práctica es considerar la creación de un plan y financiación suficiente para asegurar la viabilidad y escalamiento a largo plazo , que requiere la implementación de un repositorio de datos de investigación. Aunque el contexto sea la CA y el AA, los recursos financieros son indispensables y abarcan muchos pilares de un proyecto de esta magnitud.
4	Con el objeto de facilitar la aprobación de estos proyectos, se podrían realizar convenios entre varios estamentos públicos y privados con el propósito de distribuir los costos de implementación y operación. Si bien para las instituciones que considerarían un repositorio como un elemento diferenciador o como una ventaja estratégica en algunas áreas esto puede ser difícil de implementar, sí podría ser una forma de abordar un proyecto que, de otro modo, sería irrealizable por la falta de experiencia, conocimiento y/o recursos.
5	Una recomendación en esta categoría es, también, planificar costos económicos para las capacitaciones del equipo de proyecto o asesorías necesarias, ya que, por tratarse de un tema de diversa complejidad, y más bien nuevo para muchos países en desarrollo, esta recomendación se convierte prácticamente en una necesidad ineludible para las organizaciones.
Recursos Tecnológicos	
1	Sin lugar a dudas, en esta categoría se concentra una gran cantidad de buenas prácticas que es imperioso acoger en beneficio de la implementación de repositorios de datos de investigación. Estas buenas prácticas de índole tecnológica emanan desde las principales organizaciones de investigación en el mundo, en especial de Europa (como por ejemplo OpenAIRE y éste, a su vez, en LA Referencia en el caso de nuestro continente) y de Estados Unidos, encontrándose alineadas con el desarrollo y avance de infraestructuras tecnológicas colaborativas para propiciar y cumplir con lo que en esencia promueven los Principios FAIR.
2	Estándares – Metadatos: al momento de optar por alguno de éstos, se recomienda tener la certeza de que la elección responderá a la naturaleza de los datos almacenados en el repositorio que se va a implementar. De este modo, si el repositorio de datos es institucional, puede ser de utilidad definir un estándar de metadatos de línea más generalista; sin embargo, si el repositorio de datos de investigación es temático, probablemente será necesario seleccionar un estándar especializado que responda a la necesidad de esos investigadores. Detalles de buenas prácticas, para esta elección, se encuentran también en el contenido de esta investigación.
3	Se recomienda conocer una cosecha de datos a través de un “modelo distribuido” que debería evaluarse, considerando que la cosecha de datos permita a cada investigador, independientemente de la institución y/o país donde se encuentre, tener las mismas posibilidades de que sus datos sean cosechados directamente por un repositorio genérico, en vez de hacerlo como ocurre hoy, en general, en los países menos desarrollados. Así, al acortar los estratos de cosecha, el investigador tiene la garantía de que sus datos no se extraviarán o

	desaprovecharán al pasar por diversos filtros; al contrario, tendrán las mismas oportunidades de visibilización y uso (nuevamente, los Principios FAIR) que las de cualquier otro investigador, independiente del país o lugar de residencia.
4	Software: la elección de éste constituye una de las decisiones más relevantes al momento de implementar un repositorio de datos de investigación y su posterior gestión. Como recomendación, se debería escoger un software que cumpla con varios requisitos, tales como código abierto, facilidad de instalación, lenguaje predeterminado con opción de otros idiomas de uso, compatibilidad con el sistema informático existente, capacidad de almacenamiento y preservación de datos en el largo plazo, la colección de conjuntos de datos alojados, metadatos, herramientas de curaduría para depurar los datos de investigación, capacidad de gestionar datos según los Principios FAIR, redes de soporte adecuadas, que no requiera de muchas migraciones a versiones (deseable) y reconocida calidad demostrada a través de testimonios de casos exitosos. En este contexto, la buena práctica es informarse previamente de los softwares disponibles, considerando los aspectos anteriormente mencionados. Dentro de los más usados en la actualidad se encuentran DSpace, Dataverse, Figshare y Fedora, todos presentes en diversos países del mundo y con una trayectoria reconocida en el mercado. Para la selección de un software se recomienda, también, consultar directorios especializados de repositorios de datos de investigación, como por ejemplo re3data.org o DataCite (este último contiene registros de repositorios, incluyendo de datos, del tipo temáticos).
5	Identificadores persistentes: una buena práctica es adoptarlos para optimizar la visibilidad. En el caso específico de los repositorios de datos de investigación, se recomiendan para cada conjunto de datos. Algunos ejemplos son el identificador de objetos digitales (DOI) o el número de acceso, garantizando que una localización o ubicación determinada estarán almacenadas y preservadas en el conjunto de datos. Asimismo, en el caso de un DOI, una buena práctica es hacer referencia al conjunto de datos específico y no simplemente a una colección más amplia de datos. Actualmente, los identificadores persistentes abiertos son reconocidos, porque permiten la interoperabilidad y brindan facilidades de integración en el ecosistema digital. También se recomienda conocer ORCID y evaluar la factibilidad de incorporarlo al repositorio de datos, ya que éste es reconocido para la identificación de autores. Lo mismo se recomienda en el caso de Handle ²⁴⁵ para efectos de un sistema de identificación y recuperación de recursos y objetos digitales.
6	Otra buena práctica es considerar copias diarias de seguridad de todas las bases de datos y archivos , ante la eventualidad de una falla en el sistema, y un plan de respaldo de programas y parametrización con cierta periodicidad (dos o tres veces en un año). También es una buena práctica restringir el acceso para controlar eventuales riesgos del sistema. En dicho sentido, frente a operadores no autorizados, los datos confidenciales pueden estar expuestos si se almacenan en una única ubicación. Entonces, la recomendación es considerar y proponer soluciones para evitar este riesgo. En este mismo contexto, una buena práctica es implementar protocolos de seguridad en una única ubicación de almacenamiento de datos, en vez de hacerlo en varias. Los esquemas de alta disponibilidad -que aseguran que, cuando

²⁴⁵ <https://www.handle.net/>

	parte de los servidores falla, la operación es asumida automáticamente por servidores duplicados- también son un elemento para considerar cuando es alto el impacto de la no disponibilidad del sistema por un periodo prolongado. Por último, el uso de servicios en la nube es una forma cada vez más frecuente para reducir los costos y personal especializado respecto de la administración del hardware local, lo que también se sugiere considerar.
7	Una buena práctica es que las citas se puedan generar automáticamente desde el repositorio de datos de investigación. Una recomendación puede ser revisar R , el cual -además de corresponder a un lenguaje de programación con una orientación estadística- dispone de un comando para citas ²⁴⁶ que genera citas para paquetes.
8	La velocidad de la plataforma tecnológica puede verse afectada ante el aumento de los conjuntos de datos. Como una buena práctica, hay que asegurarse con antelación y planificar que el sistema de administración de la base de datos del repositorio pueda escalar con la expansión de datos.
9	En términos generales, una buena práctica en relación a las API (Application Programming Interface) es que permitan en un 100% la interconexión de módulos y aplicaciones, facilitando el acceso a backends y admitiendo la reutilización de servicios, para lo cual se debería especificar cómo lograrían interactuar los diferentes componentes del software. Las API deberían permitir la integración con otros sistemas institucionales de gestión y de descubrimiento. Además, el repositorio de datos de investigación debería tener la capacidad de integrarse con otros repositorios y descubridores fuera de la institución.
	Recursos humanos
1	Desde el punto de vista de recursos humanos, las recomendaciones recopiladas se orientan a que las instituciones dispongan de equipos de dedicación exclusiva para la gestión de los repositorios, destacando -además- que sean grupos interdisciplinarios, entre los que se consideren bibliotecarios e informáticos. Mientras la labor de los primeros va más allá de la gestión de datos de investigación, considerando la labor de capacitar y de asistir a los investigadores en el proceso de los planes de gestiones de datos, los segundos -en tanto- debieran abocarse sólo a la labor de implementación tecnológica, considerando la gestión del software y realizando, por ejemplo, las actualizaciones que éste requerirá a través del tiempo.
2	Si aún la biblioteca no dispone de un área de investigación, la recomendación ineludible es que el personal pueda sumarse lo antes posible, ya sea vinculándose con aspectos tecnológicos o por medio de la generación de servicios de colaboración y apoyo a la gestión de investigación. Por el contrario, si ya se tiene un área que atienda estos temas, la recomendación es estar al día y trabajar en el posicionamiento de la unidad.
3	La implementación de un repositorio de datos de investigación -y su gestión- puede llegar a ser un gran proyecto, tiempo en el cual se deberán tomar importantes decisiones. En este sentido, la recomendación es involucrar a todas las partes interesadas , dentro de éstas las personas que tienen la capacidad de tomar decisiones.

²⁴⁶ <https://www.rdocumentation.org/packages/utils/versions/3.6.2/topics/citation>

	Interfaz web
1	La recomendación y desafío es que ésta se base en la UX (Experiencia de Usuarios), por lo que al momento de evaluar un software hay que considerar aspectos de interfaz web, incluyendo navegación en computadores personales, equipos móviles inteligentes y tablets. Si hay disponibilidad de profesionales con experiencia en esta materia, se recomienda considerar aspectos de la arquitectura web para el diseño y organización de la interfaz. Actualmente, existe una serie de herramientas, tales como test de usabilidad y cardsorting. Asimismo, existe la posibilidad de realizar entrevistas, cuyos resultados podrán evaluarse, entregando directrices que dejarán satisfechos a los usuarios de su repositorio de datos de investigación.
2	Desde el inicio de la implementación de un repositorio de datos, se recomienda pensar en la organización interna y externa de la colección de datos y conjuntos de datos. También se recomienda conocer el área digital de la institución donde se implementará el repositorio, accediendo a recomendaciones relacionadas, por ejemplo, con un kit digital con el propósito de que la interfaz se desarrolle en armonía con las normas digitales institucionales.

10. Conclusiones

Implementar y gestionar un repositorio de datos de investigación, ¿es un desafío para las bibliotecas? Esta fue la pregunta de inicio que me permitió reflexionar concienzudamente y trazar una ruta para abordar los contenidos de esta tesis. Si bien existe gran cantidad de literatura, sentí que una publicación sobre este tema podría ser de utilidad para algunas bibliotecas y para quienes requiriesen conocer sobre el contexto de la Ciencia Abierta y el Acceso Abierto y su relación con los repositorios; en este caso, con repositorios de datos de investigación. Nuevas interrogantes fueron surgiendo, entre las cuales figuran: ¿existe un marco regulatorio?, ¿cómo pueden las bibliotecas implementar un repositorio?, ¿qué software elegimos?, ¿abordamos el desafío en la institución o adherimos a alguna organización que ya dispone de repositorios de datos?

Ineludibles preguntas que me incentivaron a continuar el proceso de investigación, el cual concluyó con el acopio de mucha información, incluyendo los resultados de una encuesta pionera en esta materia, que espero responda -como fue en mi situación- al planteamiento de las preguntas preliminares, que considero son relevantes de considerar para que los equipos de bibliotecas puedan emprender con la reflexión en torno a las recomendaciones y buenas prácticas más adecuadas. Asimismo, con el fin de responder satisfactoriamente en la implementación o mejoras, aparte de una adecuada gestión, de un repositorio de datos de investigación.

A continuación, comparto algunas conclusiones generales y otras más específicas, dejando abierta la posibilidad de que quienes hayan leído esta publicación -en forma completa o parcial- puedan también motivar el surgimiento de sus propias conclusiones:

- Se constata una tendencia en Europa, Australia y Estados Unidos en cuanto a promover y robustecer la Ciencia Abierta y el Acceso Abierto, amparada en la existencia de acuerdos y políticas nacionales e interregionales que tienen como fin instalar una nueva forma de abordar los resultados de la investigación. Dicha tendencia suma a esta nueva visión que se viene centrando en provocar y promover un cambio en aspectos económicos y sociales, los cuales han desempeñado un rol en esta preferencia, por medio de la que se empezó a reconocer que los descubrimientos o avances generados en ciencias y/o tecnologías -en este caso, datos de investigación-, financiados con recursos públicos, deben ser abordados adecuadamente en aspectos del tratamiento. De esta manera se aseguran, por ejemplo, la aplicación de los Principios FAIR -hoy reconocidos a nivel mundial-, directrices idóneas, estándares aprobados y certificados, jurisdicción y ética, al igual que la implementación de estructuras tecnológicas adecuadas. Todo lo anterior como parte del ciclo de elementos necesarios para garantizar el bienestar de las personas con el fin de que llegue a todos.
- Se concluye, además, que Latinoamérica -y Chile, por supuesto- se están sumando a la tendencia anteriormente descrita, porque, hacerlo, significa no sólo beneficiarse del conocimiento que se genera en los países desarrollados, sino también implica aportar a la

comunidad de investigación mundial con conocimiento generado dentro de nuestra zona geográfica, igualmente con características y atributos de calidad – y por cierto- innovador, lo que favorece la visibilidad, impacto y generación de colaboración insospechadas. Sumarse a esta tendencia significa, también, aceptar y aplicar las normativas vigentes respecto a las infraestructuras tecnológicas, a temas de gestión, jurídicos y éticos concernientes al tratamiento de datos de investigación, incluyendo su uso muchas veces definido por cada institución. En este último caso, se puede concluir que, además de ser muy pocos los investigadores que están dispuestos a compartir sus datos, hay en cambio una tendencia creciente por el interés en utilizarlos, lo que se convierte en una oportunidad en relación a la curatoría de los datos de investigación y a desarrollar tecnologías confiables, lo cual será un incentivo para los investigadores, quienes percibirán que sus datos serán gestionados y preservados adecuadamente, porque el repositorio de datos que se les ofrezca cumplirá con los estándares y normativas legales y éticas que requieren.

- Se puede concluir que, para avanzar, el aspecto de la inversión pública adquiere una preponderancia fundamental, ya que el flujo de recursos económicos propicia las condiciones necesarias para estimular el desarrollo de las investigaciones, conducentes -en definitiva- a una mejor ciencia en términos globales. Por un lado, se adquiere prestigio por el conocimiento puro y, por otro, se beneficia la sociedad al traducir dicho conocimiento en un servicio tangible que mejora la calidad de vida de las personas. Además, conlleva otro impacto aún más notable: pone el conocimiento al alcance no sólo de los investigadores, sino de cualquier persona que desee acceder a él. Sin embargo, para lograr este escenario, en términos gubernamentales se debería considerar que las políticas fuesen diseñadas invitando a conocer los beneficios sociales anteriormente esbozados, con el fin de incentivar a los investigadores a compartir los resultados de sus descubrimientos.
- Es indispensable contar con un marco regulatorio para todas las instituciones dentro de un mismo país, el cual esté en conformidad con las directrices emanadas de las organizaciones que lideran estas materias en el mundo, especialmente de los países de la Unión Europea, cuya experiencia y trayectoria están muy por delante de la realidad de países como el nuestro. Al respecto, es muy valiosa la labor desarrollada por instancias como LA Referencia o la ANID, ésta última en un proceso de retomar estas materias con renovado ahínco en nuestro país. Para lograrlo, han involucrado a todos los entes posibles por medio de una consulta pública que incluyó a los datos de investigación y cuyos resultados serán presentados, según lo publicado, durante el 2021.
- Aún hay cierta renuencia, por parte de algunos investigadores, para compartir sus datos de investigación en repositorios de Acceso Abierto, debido a la ausencia de compensaciones más allá de una cita bibliográfica en la investigación de otro colega, como podría ser una retribución económica. Asimismo, parte de esta conducta se explicaría por una creencia relacionada con los derechos de propiedad intelectual, que se asumirían como privados incluso cuando el financiamiento sea público. Este hecho ocurriría a nivel mundial, aunque

es más frecuente en Latinoamérica. El desafío en este sentido es para nuestra labor profesional, significando que podemos aportar y propiciar la comprensión en estas materias, lo cual constituye un eje para generar el cambio cultural que se requiere.

- El aspecto de infraestructura tecnológica es de vital importancia en el éxito de la implementación y en el desarrollo de los repositorios para los datos de investigación. En la actualidad, las instituciones disponen de una oferta variada de softwares, aunque la preferencia, como lo demuestran las estadísticas, está puesta en un par de ellos, luego de ser evaluados en términos de costos versus beneficios, de gestión o de preservación. La selección de este aspecto debe considerar, también, las necesidades de la organización, entendiendo por tales el nivel de producción científica que se genera, los objetivos contemplados en el plan estratégico relacionado con el área de investigación, las compatibilidades informáticas, fortalezas y debilidades del mismo, reputación y prestigio del proveedor, y servicio técnico, entre otros. Asimismo, esta tecnología debe diseñarse para facilitar aspectos conducentes a una adecuada descripción para los conjuntos de datos, permitiendo adoptar fácilmente estructuras de datos enriquecidas con las recomendaciones de, por ejemplo, DataCite. Hay que considerar el uso de identificadores persistentes, incluir al menos alguna licencia, implementar protocolos (OAI-PMH) y directrices de interoperabilidad. En esta materia, las experiencias de otras instituciones deberían orientar a las que están implementando un repositorio de datos (o desean mejorarlo), respecto de sus decisiones para garantizar el buen proceso de desarrollo y el éxito en este aspecto.
- En cuanto a los datos de investigación, se evidencia la necesidad de regular su gestión, tema en el que Europa ha tomado la vanguardia, reclutando cada vez más instituciones dentro y fuera de sus fronteras con el fin de promover el avance y el desarrollo de hallazgos y de descubrimientos en las ciencias, aumentando la transparencia y facilitando tanto la reproducción como la validación de dichos hallazgos. Con ello, se repercute positivamente en la visibilidad de una investigación, contribuyendo también a multiplicar su impacto, y se fomenta un círculo virtuoso de optimización de recursos y de colaboración en los descubrimientos de las ciencias. La implementación y las buenas prácticas en la concepción de un repositorio de datos son fundamentales para evitar la alta tasa de pérdida y/o desperdicio de millones de datos de investigación, que impactan en aspectos económicos.
- El factor recurso humano también se muestra como determinante en el desarrollo exitoso de un repositorio de datos de investigación (anexo a esto, la creación de un servicio), pues requiere movilizar este aspecto, ya que -dado el grado de tecnología y de complejidad que éste implica- la capacitación y puesta al día de todos los agentes involucrados (investigadores, desarrolladores informáticos, bibliotecarios, entre otros) surge como una necesidad ineludible.

- Tomando como apoyo las experiencias y conocimientos obtenidos a través de la metodología de estudio aplicada en este trabajo, puedo concluir también que -en cuanto a implementar y/o mejorar un repositorio de datos de investigación- los beneficios para una biblioteca son inconmensurables, situándola en una posición de prestigio y valoración por el círculo de investigación cercano, al cual le puede brindar gran apoyo y soporte. También las bibliotecas que disponen de repositorios de datos son apreciadas por estar a la vanguardia, convirtiéndose en un referente para impulsar el desarrollo profesional de la especialidad. En otro aspecto, las instituciones que apoyan la política de implementar un repositorio de datos de investigación (o mejorarlo, en caso de que una organización ya cuente con él) percibirán el valor que se le puede otorgar al plan de desarrollo estratégico, siendo un factor relevante que contribuye en el prestigio y posicionamiento institucional.
- Finalmente, puedo concluir y sugerir como una buena práctica –en lo posible de acción inmediata- que, para incrementar los datos en un repositorio, ya sea de publicaciones o mixto (publicaciones y datos), se recomienda tomar las publicaciones ya almacenadas en su repositorio y evaluar si éstas contienen datos asociados o complementarios, los cuales pueden ser tratados como otros registros, aplicando los metadatos, entre otras medidas. De esta forma, se incrementa su colección de datos institucional y se hacen más visibles, considerando su vínculo tanto en las publicaciones como en los datos de investigación.

Bibliografía consultada

- Amaro, B. S. (04 de Septiembre de 2020). *TICAL2020 - Plenaria 4 (final): Manejo de datos abiertos para potenciar la ciencia en la Universidad*. Obtenido de TICAL2020 - Plenaria 4 (final): Manejo de datos abiertos para potenciar la ciencia en la Universidad: https://www.youtube.com/watch?v=o2MSVkbRdWo&list=PLsPfupT39gzO7l1xMfefa_scmlhvpCXTN&index=6
- AmeliCA. (s.f.). *AmeliCA*. Obtenido de AmeliCA. Conocimiento Abierto sin fines de lucro propiedad de la Academia: <http://amelica.org/>
- ANID. Agencia Nacional de Investigación y Desarrollo. (Junio de 2020). *Propuesta de Política de acceso abierto a la información científica y a datos de investigación financiados con fondos públicos de la ANID*. Obtenido de ANID. Agencia Nacional de Investigación y Desarrollo Web: https://s3.amazonaws.com/documentos.anid.cl/estudios/Politica_acceso_a_informacion_cientifica_version_final_26-05-2020.pdf
- ANID. Agencia Nacional de Investigación y Desarrollo. (Noviembre de 2020). *Resultados de la Consulta Pública Política de Acceso Abierto a Información Científica y a Datos de Investigación*. Obtenido de Resultados de la Consulta Pública Política de Acceso Abierto a Información Científica y a Datos de Investigación: https://www.uahurtado.cl/wp-images/uploads/2020/11/ResultadosConsultaPublica_PAA_2020.pdf
- ANOVO. (8 de Noviembre de 2019). *¿Qué es un agregador y qué tipos existen?* Obtenido de ANOVO : <https://www.anovo.es/que-es-un-agregador/>
- ARDC. Australian Research Data Commons. (s.f.). *Cita e identificadores*. Obtenido de Cita e identificadores: <https://ardc.edu.au/resources/working-with-data/citation-identifiers/>
- Baker, K. D. (2017). *Research and the Changing Nature of Data Repositories (pp. 33-60)*. Obtenido de Curating Research Data. Practical Strategies four your Digital Repository. Volume One: http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/booksanddigitalresources/digital/9780838988596_crd_v1_OA.pdf
- CE. Comunidad Europea. (s.f.). *¿Qué es el Acceso Abierto?* Obtenido de Horizon 20202 Online Manual : https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm
- CE. Comunidad Europea. (2017). *Guías Horizonte 2020*.
- CERN y colaboradores. (2016-2019). *Gobernancia*. Obtenido de Invenio Open Science Web Site: <https://invenio-software.org/governance/>
- COAR. Building a Global Knowledge Commons. (28 de Noviembre de 2017). *Next Generation Repositories. Behaviours and Technical Recommendations of the COAR. Next Generation*

- Repositories Working Group*. Obtenido de NGR Final Formatted Report: <https://www.coar-repositories.org/files/NGR-Final-Formatted-Report-cc.pdf>
- COAR. Confederation Open Access Repositories. (s.f). *About the Technologies*. Obtenido de Next Generation Repositories: <https://ngr.coar-repositories.org/technology/>
- COAR. Confederation Open Access Repositories. (s.f). *Defining the next generation repository*. Obtenido de Next Generation Repositories: <https://ngr.coar-repositories.org/>
- COPDESS. Coalition for Publishing Data in the Earth and Space Sciences. (s.f.). *Declaración de Compromiso en las Ciencias de la Tierra, el Espacio y el Medio Ambiente*. Obtenido de Declaración de Compromiso en las Ciencias de la Tierra, el Espacio y el Medio Ambiente: <https://copdess.org/enabling-fair-data-project/commitment-statement-in-the-earth-space-and-environmental-sciences/>
- Cousijn, H. D. (29 de Mayo de 2020). *The why, what, and how of repositories*. Obtenido de DataCite Blog: <https://blog.datacite.org/the-why-what-and-how-of-repositories/>
- Cousijn, H. y. (6 de Febrero de 2020). *German Research Foundation to Fund New Services of re3data*. Obtenido de DataCite Blog: <https://blog.datacite.org/german-research-foundation-to-fund-new-services-of-re3data/>
- Creative Commons. (20 de Junio de 2020). *Atribución 4.0 Internacional (CC BY 4.0)*. Obtenido de Creative Commons : <https://creativecommons.org/licenses/by/4.0/>
- Creative Commons. (20 de Junio de 2020). *CC0 1.0 Universal (CC0 1.0) Public Domain Dedication*. Obtenido de Creative Commons : <https://creativecommons.org/publicdomain/zero/1.0/>
- CRL. Center for Research Libraries Global Resources Network. (s.f). *Certification and Assessment of Digital Repositories*. Obtenido de CRL. Center for Research Libraries Global Resources Network: <https://www.crl.edu/archiving-preservation/digital-archives/certification-assessment>
- CRL. Center for Research Libraries. Global Resources Network. (s.f). *Ten Principles*. Obtenido de CRL (Center for Research Libraries. Global Resources Network): <https://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/core-re>
- Dataverse Project. (2020). *The Strategic Goals*. Obtenido de Dataverse Project: <https://dataverse.org/about>
- DCMI. Dublin Core [™] Metadata Initiative Web site. (1995-2020). *Guía del usuario de Dublin Core [™]*. Obtenido de DCMI. Dublin Core [™] Metadata Initiative Web site: <http://dublincore.org/resources/userguide/> Consultado el 27 de Junio de 2020
- DCMI. The Dublin Core[™] Metadata Initiative. (1995-2020). *Conceptos básicos de metadatos*. Obtenido de DCMI. The Dublin Core[™] Metadata Initiative Web site: <http://dublincore.org/resources/metadata-basics/> Consultado 27 de Junio de 2020

- De Filippo, D., & y D'Onofrio, M. (Noviembre de 2019). *Alcances y limitaciones de la ciencia abierta en Latinoamérica: análisis de las políticas públicas y publicaciones científicas de la región*. Obtenido de Hipertext.net (N°19, pp. 32-48):
<https://www.raco.cat/index.php/Hipertext/article/view/360106>
- Desarrollo, A. A. (06 de Agosto de 2020). *Con éxito concluye política sobre Acceso Abierto a Datos Científicos*. Obtenido de Con éxito concluye política sobre Acceso Abierto a Datos Científicos: <https://www.anid.cl/2020/07/06/con-exito-concluye-primera-consulta-publica-sobre-acceso-abierto-a-datos-cientificos/>
- DFG. Fundación Alemana de Investigación. (6 de Febrero de 2020). *Blog de DataCite*. Obtenido de re3data.org - Registro de repositorios de datos de investigación Web site: re3data.org - Registro de repositorios de datos de investigación. <https://doi.org/10.17616/R3D> último acceso: 2020-06-21
- Donohue, T. y. (s.f.). *DSpace-CRIS Home*. Obtenido de LYRASIS: Recuperado el 20 Julio 2020 de <https://wiki.lyrasis.org/display/DSPACECRIS/DSpace-CRIS+Home>
- DSpace. (2020). *DSpace características*. Obtenido de DSpace Web site: <https://duraspace.org/dspace/>
- Eagar, M. (04 de Noviembre de 2017). *¿Cuál es la diferencia entre sistemas descentralizados y distribuidos?* Obtenido de ¿Cuál es la diferencia entre sistemas descentralizados y distribuidos?: <https://medium.com/distributed-economy/what-is-the-difference-between-decentralized-and-distributed-systems-f4190a5c6462>
- EC. European Comission. (s.f.). *Dar forma al futuro digital de Europa*. Obtenido de Dar forma al futuro digital de Europa: <https://ec.europa.eu/digital-single-market/>
- EC. European Comission. (s.f.). *Espacio Europeo de Investigación (EEI)*. Obtenido de Espacio Europeo de Investigación (EEI): https://ec.europa.eu/info/research-and-innovation/strategy/era_es
- ERC. European Research Council. (2017). *Directrices sobre la aplicación de acceso abierto a publicaciones científicas y datos de investigación en proyectos apoyados por el Consejo Europeo de Investigación (Horizonte 2020)*.
- ERC. European Research Council. (21 de Abril de 2017). *Guidelines on the Implementation of Open Access to Scientific Publications and Research Data in projects supported by the European Research Council under Horizon 2020*. Obtenido de ERC (European Research Council): https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide_en.pdf
- EU. European Union. (2012). *Portal de Datos Abiertos de la Unión Europea. Acceso a los Datos Abiertos de la UE*. Obtenido de Portal de Datos Abiertos de la Unión Europea: <https://data.europa.eu/euodp/es/about>

- EU. European Union. (20 de Junio de 2020). *2011/833/EU: Commission Decision of 12 December 2011 on the reuse of Commission documents*. Obtenido de EUR-Lex. Access to European Union Law: <https://eur-lex.europa.eu/eli/dec/2011/833/oj>
- EU. European Union. (20 de Junio de 2020). *Exención de Responsabilidad*. Obtenido de Portal de Datos Abiertos de la UE: <https://data.europa.eu/euodp/es/disclaimer>
- EU. European Union. (20 de Junio de 2020). *Página de Aviso de Derechos de Autor*. Obtenido de Portal de Datos Abiertos de la UE: <https://data.europa.eu/euodp/es/copyright>
- EU. European Union. (20 de Junio de 2020). *Portal de datos abiertos de la UE. Acceso a los Datos Abiertos de la Unión Europea*. Obtenido de Portal de Datos Abiertos de la UE: <https://data.europa.eu/euodp/es/about>
- EU. European Union. (20 de Junio de 2020). *Protección de los Datos Personales*. Obtenido de Portal de Datos Abiertos de la UE: <https://data.europa.eu/euodp/es/privacystatement>
- European Commission. (20 de Junio de 2020). *Register of the Data Protection Officer (DPO)*. Obtenido de European Commission DPO Register: <https://ec.europa.eu/dpo-register/>
- Fedora. (2020). *Fedora. Fedora is the flexible, modular, open source repository platform with native linked data support*. Obtenido de Fedora : <https://duraspace.org/fedora/>
- Fernández Eguiarte, A. y. (2-4 de Septiembre de 2019). *3er Encuentro Latinoamericano de e-Ciencia. Repositorio Institucional del Centro de Ciencias de la Atmósfera, UNAM*. Obtenido de 3er Encuentro Latinoamericano de e-Ciencia. Repositorio Institucional del Centro de Ciencias de la Atmósfera, UNAM: <http://atlasclimatico.unam.mx/presentacion.pdf>
- Figshare. (24 de Julio de 2018). *Figshare as a data repository*. Obtenido de Figshare as a data repository: https://figshare.com/blog/figshare_as_a_data_repository/410
- Figshare. (2020). *Figshare for institutions*. Obtenido de Figshare for institutions: <https://knowledge.figshare.com/institutions>
- Figshare. (2020). *Figshare.com*. Obtenido de Figshare.com: <https://figshare.com/>
- Figshare. (s.f.). *Figshare / Mark Hahnel profile*. Obtenido de Figshare / Mark Hahnel profile: https://figshare.com/authors/Mark_Hahnel/96371
- GitHub. (s.f.). *Metadatos*. Obtenido de Metadatos: <http://rd-alliance.github.io/metadata-directory/standards/>
- GOFAIR. (15 de Octubre de 2020). *Noticias / Nueva Publicación de la UE: Seis recomendaciones para la implementación de la práctica FAIR*. Obtenido de Noticias / Nueva Publicación de la UE: Seis recomendaciones para la implementación de la práctica FAIR: <https://www.go-fair.org/2020/10/30/new-eu-publication-six-recommendations-for-implementation-of-fair-practice/>
- Hahnel, M. M. (01 de Diciembre de 2020). *The State of Open Data 2020 - Digital Science Report*. Obtenido de The State of Open Data 2020 - Digital Science Report:

- https://digitalscience.figshare.com/articles/report/The_State_of_Open_Data_2020/13227875/2
- INVENIO. (© 2016-2019). *Invenio. Powering Open Science*. Obtenido de INVENIO : <https://invenio-software.org/>
- Kindling, M. P. (Marzo, Abril de 2017). *El panorama de los repositorios de datos de investigación en 2015: un análisis re3data*. Obtenido de D-Lib Magazine: <http://www.dlib.org/dlib/march17/03contents.html>
- LA Referencia. (s.f.). *LA Referencia. Red de Repositorios de Acceso Abierto a la Ciencia*. Obtenido de LA Referencia. Red de Repositorios de Acceso Abierto a la Ciencia: <http://www.lareferencia.info/es/>
- LA Referencia. Red de Repositorios de Acceso Abierto a la Ciencia. (23 de Junio de 2020). *Technology. We Develop Technology for National Harvesters*. Obtenido de LA Referencia: <http://www.lareferencia.info/en/services/tecnologia>
- Lin, D. (20 de Abril de 2017). *A Primer on the Certifications of a Trusted Digital Repository (TDR)*. Obtenido de DataScience@NHI Blog: https://datascience.nih.gov/trusted_digital_repository
- Lin, D. (s.f.). *A Primer on the Certifications of a Trusted Digital Repository (TDR)*. Obtenido de DataScience@NIH Blog: Recuperado el 26 Marzo 2019 de https://datascience.nih.gov/trusted_digital_repository
- López, W. (Abril de 2020). *La Coalición S y el Plan S: Implicaciones para los ecosistemas de conocimiento en América Latina*. Obtenido de Pontificia Universidad Javeriana-Editorial Pontificia Universidad Javeriana: <https://revistas.javeriana.edu.co/index.php/revPsycho/article/view/29651>
- Melero, R. y. (2018). *Maredata, una iniciativa para el flujo de datos compartidos*. Obtenido de Revista ORL (Volumen 9, N° 4): <https://revistas.usal.es/index.php/2444-7986/article/view/orl.16988/0>.
- Ministerio de Asuntos Económicos y Transformación Digital. Gobierno de España. (18 de 12 de 2017). *Búsqueda de repositorios de datos de investigación: re3data.org*. Obtenido de Datos.Gob.Es: <https://datos.gob.es/es/blog/busqueda-de-repositorios-de-datos-de-investigacion-re3dataorg>
- Nature. (05 de Junio de 2019). *Hacer que los datos científicos sean justos*. Obtenido de Hacer que los datos científicos sean justos: <https://www.nature.com/articles/d41586-019-01720-7?sf213819360=1>
- Natusfera. (5 de Octubre de 2017). *Conceptos de Ciencia Ciudadana*. Obtenido de Natusfera. La biodiversidad al alcance de los ciudadanos: https://www.gbif.es/wp-content/uploads/2017/12/02_Conceptos_de_CienciaCiudadana.pdf

- NIH. National Institutes of Health. (s.f.). *Data Infrastructure*. Obtenido de Office of Data Science Strategy: Recuperado 10 Diciembre 2019 de <https://datascience.nih.gov/data-infrastructure>
- NIH. National Institutes of Health. (s.f.). *NIH Sharing Policies and Related Guidance on NIH-Funded Research Resources*. Obtenido de NIH. Grants and Funding: Recuperado el 7 Febrero 2018 de <https://grants.nih.gov/policy/sharing.htm>
- NIH. National Institutes of Health. (s.f.). *NIH Strategic Plan for Data Science*. Obtenido de Office of Data Science Strategy: Recuperado 14 Septiembre 2020 de <https://datascience.nih.gov/strategicplan>
- Obama, B. (9 de Mayo de 2013). *Open Government Initiative*. Obtenido de The White House : <https://obamawhitehouse.archives.gov/open>
- OKF. Open Knowledge Foundation. (2020). *CKAN, The World's Leading Open Source Data Portal Platform*. Obtenido de CKAN (Comprehensive Knowledge Archive Network): <https://ckan.org/>
- OpenAIRE. (2015). *Directrices de OpenAIRE para Archivos de Datos*. Obtenido de Directrices de OpenAIRE para Archivos de Datos: <https://guidelines.openaire.eu/en/latest/data/index.html>
- OpenAIRE. (29 de Abril de 2018). *OpenAIRE History*. Obtenido de OpenAIRE : <https://www.openaire.eu/history>
- OpenAIRE. (4 de Enero de 2019). *OpenAIRE Guidelines for Literature Repository Managers v 4.0 are now available!* Obtenido de OpenAIRE Guidelines for Literature Repository Managers v 4.0 are now available! Web site: <https://www.openaire.eu/openaire-guidelines-for-literature-repository-managers-v-4-0-are-now-available> Consultado el 27/06/20
- OpenAIRE. (20 de Junio de 2020). *Conoce a OpenAIRE nuestras fases*. Obtenido de OpenAIRE Web site: <https://www.openaire.eu/openaire-history>
- OpenAIRE. (s.f.). *Ciencia Abierta en Europa*. Obtenido de Ciencia Abierta en Europa: <https://www.openaire.eu/open-science-europe-overview>
- OpenAIRE. (s.f.). *Cómo encontrar un repositorio confiable para sus datos / Guía para investigadores*. Obtenido de Cómo encontrar un repositorio confiable para sus datos / Guía para investigadores: <https://www.openaire.eu/find-trustworthy-data-repository>
- OpenAIRE. (s.f.). *Conoce OpenAIRE*. Obtenido de Conoce OpenAIRE: <https://www.openaire.eu/openaire-history>
- OpenAIRE. (s.f.). *OpenAIRE*. Obtenido de OpenAIRE: <https://www.openaire.eu/>
- OpenAIRE. (s.f.). *OpenAIRE - What is the Open Research Data Pilot?* Obtenido de OpenAIRE - What is the Open Research Data Pilot?: <https://www.openaire.eu/what-is-the-open-research-data-pilot>

- OpenAIRE. (s.f.). *OpenAIRE en EOSC: dónde contribuimos*. Obtenido de OpenAIRE en EOSC: dónde contribuimos: <https://www.openaire.eu/openaire-and-eosc>
- PLOS ONE Editors. (1 de Marzo de 2018). *PLOS Criteria for Recommended Data Repositories*. Obtenido de EveryONE PLOS: <https://everyone.plos.org/2018/03/01/criteria-for-recommended-data-repositories/>
- re3data.org. (10 de 06 de 2020). *Registro de repositorios de datos de investigación*. Obtenido de Registro de repositorios de datos de investigación.: <https://doi.org/10.17616/R3D>
- Rufino Campêlo, L. y. (20 de Marzo de 2020). *Comparación de software gratuito para la creación de repositorios de datos abiertos*. *Ciencias de la información* , 48 (3). Obtenido de Comparación de software gratuito para la creación de repositorios de datos abiertos. *Ciencias de la información* , 48 (3): <http://revista.ibict.br/ciinf/article/view/5004>
- Schmidt, B. y. (20 de Junio de 2020). *Implementing Open Access Mandates in Europe: OpenAIRE Study on the Development of Open Access of Repository Communities in Europe (p.11)*. Obtenido de Universitätsverlag Gottingen. OpenAIRE: https://univerlag.uni-gottingen.de/bitstream/handle/3/isbn-978-3-86395-095-8/oa_mandates.pdf?sequence=4&
- SciELO . (16 de Marzo de 2016). *Principios Rectores FAIR Publicados en una Revista del Nature Publishing Group*. Obtenido de Principios Rectores FAIR Publicados en una Revista del Nature Publishing Group: <https://blog.scielo.org/es/2016/03/16/principios-rectores-fair-publicados-en-una-revista-del-nature-publishing-group/#.X90kidhKjDc>
- SciELO. Scientific Electronic Library Online. (s.f.). *Declaración de Acceso Abierto*. Obtenido de SciELO. Scientific Electronic Library Online : Recuperado el 01 Septiembre 2020 de <https://scielo.org/es/sobre-el-scielo/declaracion-de-acceso-abierto/>
- Spichtinger, D. (s.f.). *Datos de Investigación Abiertos / FAIR en Horizonte 2020*. Obtenido de Datos de Investigación Abiertos / FAIR en Horizonte 2020: https://ec.europa.eu/easme/sites/easme-site/files/open_fair_research_data_in_h2020.pdf
- SpringerNature. (2020). *Recommended Data Repositories*. Obtenido de Scientific Data : <https://www.nature.com/sdata/policies/repositories>
- UAEM. Universidad Autónoma del Estado de México. (2020). *¿Qué es Redalyc.org?* Obtenido de Redalyc.org: <https://www.redalyc.org/redalyc/acerca-de/mision.html>
- UE. Unión Europea. (20 de Junio de 2020). *Portal de datos abiertos de la UE*. Obtenido de Política de cookies Web site: <https://data.europa.eu/euodp/es/cookiepolicy>
- U-GOB. Tecnología en Gobierno. (2019). *REconocimiento al Proyecto de Datos Abiertos 2019 / Categoría Organismo Autónomo Federal*. Obtenido de REconocimiento al Proyecto de Datos Abiertos 2019 / Categoría Organismo Autónomo Federal: http://atlasclimatico.unam.mx/uniatmos_portal/noticias/RepositorioInstitucional.html

- UNAM. Universidad Nacional Autónoma de México. (2017). *Centro de Ciencias de la Atmósfera / Directorio*. Obtenido de Centro de Ciencias de la Atmósfera / Directorio:
<https://www.atmosfera.unam.mx/category/agustin-fernandez-eguiarte/>
- UNAM. Universidad Nacional Autónoma de México. (2018). *Latindex. Sistema Regional de Información en Línea para Revistas Científicas de América Latina, el Caribe, España y Portugal*. Obtenido de Latindex: <https://www.latindex.org/latindex/descripcion>
- Wilkinson, M. D. (15 de Marzo de 2016). *The FAIR Guiding Principles for scientific data management and stewardship*. Obtenido de Scientific Data:
<https://www.nature.com/articles/sdata201618>"

Anexo 1: Exención de Responsabilidad

Cabe mencionar que, en el aviso jurídico de Exención, se encuentra publicada una “exención de responsabilidad” habilitada ante la salida del Reino Unido de la UE. La Oficina de Publicaciones de la UE (con sede en Luxemburgo) está actualizando parte del contenido de esta web, en vista del retiro de Reino Unido (EU. European Union, 2020).

En su propósito de orientar el portal de datos abiertos a los países e instituciones participantes, en lo referente a derechos de autor, la UE estipuló el uso de licencias que se reiteran en el ámbito de los repositorios de datos.

En su página de avisos de derechos de autor UE 2012-2020, reitera que la reutilización se rige por la mencionada anteriormente Decisión de la Comisión, de 12 de diciembre de 2011, relativa a la reutilización de los documentos de la Comisión (EU. European Union, 2020)²⁴⁷.

	Por lo que este sitio web se publica bajo la licencia Creative Commons Atribución 4.0 Internacional (CC BY 4.0) (Creative Commons, 2020)
---	--

Es posible que se deba obtener derechos adicionales si un contenido concreto representa a particulares identificables o incluye obras de terceros. Para utilizar o reproducir contenidos ajenos a la UE, puede ser necesario conseguir la autorización de los titulares de los derechos. Los programas informáticos y los documentos sometidos a derechos de propiedad industrial (como patentes y marcas registradas, dibujos o modelos registrados, logotipos y nombres) quedan excluidos de la política de reutilización de la Comisión, por lo que no se autoriza su uso (EU. European Union, 2020) (Creative Commons, 2020).

En relación a los metadatos del portal de datos abiertos, se declaran de dominio público con arreglo al resumen de Declaración de Dominio Público CCO 1.0 Universal de Creative Commons²⁴⁸.

²⁴⁷ A menos que se indique lo contrario (por ejemplo, en los avisos específicos de derechos de autor), el contenido editorial propiedad de la Unión Europea mostrado en este sitio web se publica bajo la licencia Creative Commons Atribución 4.0 Internacional (CC BY 4.0). Por consiguiente, se permite la reutilización siempre que la fuente esté adecuadamente identificada y se indique cualquier cambio realizado. Sitio web: <https://creativecommons.org/licenses/by/4.0/>

²⁴⁸ Sitio web: <https://creativecommons.org/publicdomain/zero/1.0/>



Para todos los recursos publicados, existe una referencia específica a la licencia con arreglo a la cual el propietario opta por publicar el recurso correspondiente. (Creative Commons, 2020)

Por otro lado, es importante destacar que el portal de datos abiertos aborda también la protección a los datos personales, por medio de una declaración de confidencialidad para tratarlos y protegerlos (EU. European Union, 2020).

Asimismo, este aviso jurídico se inspira en el nuevo Reglamento de protección de datos (UE) 2018/1725 para instituciones, organismos y agencias de la UE, entrando en vigencia el 11/12/2018 (derogando el Reglamento (CE) 45/2001). El artículo 31 de este nuevo reglamento requiere que los controladores de datos mantengan un 'registro' de cada operación de procesamiento bajo su responsabilidad y, por lo tanto, reemplaza el sistema anterior de 'notificaciones'. Durante un período transitorio, hasta que una notificación existente se haya convertido en un registro, permanecerá accesible en el registro como 'notificación heredada' (ID de referencia: DPO-XXXX) (European Commission, 2020).

Con la regularización de este relevante ámbito que debe considerarse al momento de implementar un repositorio de datos abiertos, se está respondiendo a la pregunta: ¿Por qué y cómo tratar este tipo de datos? Por cierto, estos datos emergen también en el campo investigativo, acercan a los aspectos jurídicos (en este caso, de los integrantes de la UE), determinando qué datos personales se puede recoger y/o cómo tratarlos, cómo conservar este tipo de datos, protegerlos y salvaguardarlos al tratarse de información sensible, al mismo tiempo que conocer los derechos y cómo ejecutarlos si alguna situación lo amerita. Más información, revisar nota²⁴⁹.

²⁴⁹ El delegado de protección de datos de la Comisión publica el registro de todas las operaciones de tratamiento de datos personales realizadas por la Comisión, que se han documentado y que le han sido notificadas. Se puede acceder al registro a través del siguiente enlace: <https://ec.europa.eu/dpo-register>.

La presente operación específica de tratamiento consta en el registro público del responsable de protección de datos con la siguiente referencia: DPR-EC-00458. Sitio web: <https://ec.europa.eu/dpo-register/>

Anexo 2: Formulario de Divulgación de Conjunto de Datos

1.- Información del contacto

¿Es usted el investigador principal / investigador principal? *

Sí/No

Sí es un estudiante, seleccione "No" y proporcione el nombre del miembro de la facultad con el que trabaja. Esto es especialmente importante para la investigación financiada.

Su Departamento / Unidad

2.- Información del investigador principal / colaborador

Colaboradores adicionales

Enumere todos los contribuyentes adicionales que participaron en la creación del conjunto de datos, tanto internos como externos a ISU.

3.- Información del conjunto de datos

Título de los datos

El título del conjunto de datos; no es necesario que coincida con el título de las publicaciones relacionadas.

Descripción *

Proporcione una descripción breve y no confidencial de los datos que planea publicar, incluidos temas, contenido, formato, código, tamaño, etc.

Elija una licencia de publicación para sus datos: *

NECESARIO. La licencia elegida rige lo que otras personas pueden hacer con los datos. Puede encontrar más información sobre estas licencias en: <https://instr.iastate.libguides.com/datashare/licenses>

____ CC-BY (Atribución Creative Commons)

____ CC-0 (Creative Commons Zero)

4.- ¿Los datos contienen información sobre temas sensibles como:

Información de identificación personal (PII) *

(p. ej., nombre completo, números de seguro social, número de licencia de conducir, correo electrónico o dirección postal)

Sí/No

Otra información personal que podría ser identificable *

(ej. : edad, ubicación, género, ocupaciones, etc.)

Sí/No

Especies y / o lugares protegidos. *

(p. ej., información que revelaría la ubicación de lugares o cosas que deberían protegerse)

Sí/No

Datos confidenciales o patentados *

Datos que no se pueden compartir con el público porque contienen información técnica (por ejemplo, secretos comerciales) o personal (por ejemplo, números de tarjetas de crédito).

Sí/No

Información Adicional

Incluya cualquier información adicional para ayudar a informar su divulgación, es decir, fechas límite, MTA, NDA, etc.

5.- PATROCINIO Y APOYO

¿Este conjunto de datos fue financiado (en su totalidad o en parte) por Ames Laboratory? *

Sí/No

¿Se utilizaron fuentes de financiación internas (distintas de las de sueldos) o externas en la creación de este conjunto de datos? *

Sí/No

6.- GARANTÍAS DE INVESTIGACIÓN

GARANTÍAS DE INVESTIGACIÓN: Sujetos humanos *

¿El proyecto involucró a sujetos humanos?

Sí/No

GARANTÍAS DE INVESTIGACIÓN: animales vivos *

Sí/No

¿El proyecto involucró animales vertebrados vivos?

Sí/No

GARANTÍAS DE INVESTIGACIÓN: Riesgos biológicos *

¿El proyecto involucró material biopeligroso o moléculas Sintéticas de ácido nucleico?

Sí/No

GARANTÍAS DE INVESTIGACIÓN: Agentes seleccionados *

¿El conjunto de datos incluye información sobre agentes seleccionados?

Sí/No

GARANTÍAS DE INVESTIGACIÓN: Control de exportaciones *

¿Existe un Plan de Control de Tecnología de ISU para esta investigación?

Sí/no

GARANTÍAS DE INVESTIGACIÓN: Información no clasificada controlada *

¿Se ha considerado información no clasificada controlada (CUI) algún dato?

Sí/No

Tomado de: <https://app.smartsheet.com/b/form/de43938cc6f34868930c619d568e2dca>

Anexo 3: Encuesta - Repositorios de Datos de Investigación en Bibliotecas

La encuesta le tomará **5 minutos** y tiene por objetivo conocer sus recomendaciones o buenas prácticas, dirigidas a la implementación de repositorios de datos de investigación en bibliotecas.

Sus respuestas serán anónimas y se mantendrá la confidencialidad. Hasta el 12 de noviembre de 2020 estará disponible la encuesta.
¡Gracias por su valiosa colaboración!

Survey - Repositories of Research Data in Libraries

The survey will take **5 minutes** and its objective is to know your recommendations or good practices, aimed at the implementation of repositories of research data in libraries.

Your responses will be anonymous and confidentiality will be maintained. Until November 12, 2020, the survey will be available.

Thanks for your valuable collaboration!

1. ¿Cuál es la dotación de personas que trabajan exclusivamente en el Repositorio de Datos de Investigación en su institución? / **What is the staffing of people who work exclusively on the Research Data Repository at your institution?**
 - Menor de 5 / **Less than 5**
 - Entre 5 y 10 / **Between 5 and 10**
 - Más de 10 / **More than 10**
2. ¿Cuál es el nombre del software con el cual implementó su repositorio? / **What is the name of the software that you used to implement your data repository?**
3. ¿Qué elemento fue clave en la selección del software (3 max.)? / **What elements were important in the software selection (max. 3)?**
 - El costo total de la implementación fue menor que otros / **The total cost of the implementation was lower than the others?**
 - La funcionalidad se adecuó a lo requerido / **The functionality was according to the requested?**

- El proveedor es reconocido en el mercado / **The supplier is well known in the industry.**
 - Experiencia previa con el software / **Previous experience with the software**
 - Buen soporte post implementación / **Good support after implementation**
 - Fue impuesto por la organización / **It was imposed by the organization**
4. El repositorio de su institución, ¿es exclusivamente para datos? / **Is the repository of your organization exclusively for data?**
- Sí / **Yes**
 - No / **Not**
5. ¿Tiene el repositorio de su institución una sección para datos de investigación, entre otras secciones? / **Does your organization's repository have a research data section, among other sections?**
- Sí / **Yes**
 - No / **Not**
6. ¿Hace cuánto tiempo el repositorio de su organización fue implementado para datos de investigación? / **How long has your institution's data repository implemented for research data?**
- Menos de cinco años / **Less than five years**
 - Cinco años / **Five years**
 - Más de cinco años / **More than five years**
7. ¿Con qué tipo de financiamiento cuenta el repositorio y/o la sección de datos de investigación? / **What type of financing does the data repository and/or the research data section have?**
- Institucional o de la organización / **Institutional or organizational**
 - Gubernamental / **Governmental**
 - Adjudicación de proyectos / **Project award**
 - Sin financiamiento / **No financing**

8. ¿Cuán satisfechos están actualmente con el software del repositorio? / **How satisfied are you currently with the data repository software?**

- Extremadamente satisfechos / **Extremely satisfied**
- Muy satisfechos / **Very satisfied**
- Algo satisfechos / **Somewhat satisfied**
- No tan satisfechos. ¿Por qué? / **Not so satisfied. Why?**
- Nada satisfechos ¿Por qué? / **Not satisfied. Why?**

9. ¿Por qué su respuesta en la pregunta anterior es "No tan satisfecho" o "Nada Satisfecho"? / **Why your answer in the previous question is "Somewhat satisfied" or "Not satisfied"?**

(Antonio esto lo agregué yo para explicarte que es una pregunta abierta y sólo se desplegó si contestaba “No tan satisfechos” o “Nada satisfechos”)

10. ¿Cuán satisfechos están con la seguridad de los datos de investigación almacenados en el repositorio? / **How satisfied are you with the security of the research data stored in the repository?**

- Extremadamente satisfechos / **Extremely satisfied**
- Muy satisfechos / **Very satisfied**
- Algo satisfechos / **Somewhat satisfied**
- No tan satisfechos. ¿Por qué? / **Not so satisfied. Why?**
- Nada satisfechos ¿Por qué? / **Not satisfied. Why?**

11. ¿Por qué su respuesta en la pregunta anterior es "No tan satisfechos" o "Nada Satisfechos"? / **Why your answer in the previous question is "Somewhat satisfied" or "Not satisfied"?**

(Antonio esto lo agregué yo para explicarte que es una pregunta abierta y sólo se desplegó si contestaba “No tan satisfechos” o “Nada satisfechos”)

12. ¿Cuán satisfechos están con la confiabilidad en relación a criterios de auditoría y certificación del repositorio de datos? / **How satisfied are you with the reliability in relation to audit criteria and data repository certification?**

- Extremadamente satisfechos / **Extremely satisfied**
- Muy satisfechos / **Very satisfied**
- Algo satisfechos / **Somewhat satisfied**
- No tan satisfechos. ¿Por qué? / **Not so satisfied. Why?**
- Nada satisfechos ¿Por qué? / **Not satisfied. Why?**

13. ¿Por qué su respuesta en la pregunta anterior es "No tan satisfecho" o "Nada Satisfecho"? / **Why your answer in the previous question is "Somewhat satisfied" or "Not satisfied"?**

(Antonio esto lo agregué yo para explicarte que es una pregunta abierta y sólo se desplegó si contestaba "No tan satisfechos" o "Nada satisfechos")

14. Indique los aspectos más complejos que tuvieron que abordar durante la implementación del software o sección para datos de investigación / **Would you like to share briefly the most complex aspects in the implementation of the software and the research data section?**

15. Comparta brevemente las principales recomendaciones (o buenas prácticas) para las bibliotecas que están implementando un repositorio o una sección de datos para investigación / **Would you like to share briefly the main recommendations (or good practices) for libraries that are implementing a data repository or a research data section**

¡Muchas gracias! / **Thank you very much!**

Amelia Muñoz Bravo

Magíster en Procesamiento y Gestión de la Información ©
Pontificia Universidad Católica de Chile