



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

**POSGRADO EN BIBLIOTECOLOGÍA Y ESTUDIOS DE LA INFORMACIÓN
FACULTAD DE FILOSOFÍA Y LETRAS
INSTITUTO DE INVESTIGACIONES BIBLIOTECOLÓGICAS
Y DE LA INFORMACIÓN**

**Desarrollo de un Repositorio de datos científicos de apoyo
a la investigación: el caso de las ciencias de la tierra**

T E S I S
**QUE PARA OPTAR POR EL GRADO DE
MAESTRA EN BIBLIOTECOLOGÍA Y ESTUDIOS DE LA INFORMACIÓN**

PRESENTA:

MINERVA CASTRO ESCAMILLA

**Director de Tesis:
Dr. Juan Voutssás Márquez
Instituto de Investigaciones Bibliotecológicas y Estudios de la Información
UNAM**

Ciudad Universitaria, Cd Mx., agosto 2020



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Agradecimientos

Mi agradecimiento a la Universidad Nacional Autónoma de México por formar parte de la comunidad académica, máxima casa de estudios.

A mi asesor Juan Voutssás Márquez, por el apoyo y pláticas que mantuvimos en todo este proceso.

También tienen mi agradecimiento mis revisores:
Dr. Jonathan Hernández Pérez,
Dr. Jaime Ríos Ortega,
Dr. César Augusto Ramírez Velázquez y al
Dr. Filiberto Felipe Martínez Arellano
Por esa revisión tan minuciosa que realizaron al trabajo y todas esas conversaciones, reuniones e intercambios de ideas que tuvimos en este proceso, sin duda mejoró éste trabajo profesional.

A la Biblioteca Conjunta de Ciencias de la Tierra (BCCT), que a través de ella he podido llevar a cabo tanto la teoría como la práctica de lo que implica la Bibliotecología y Estudios de la Información.

A mis compañeros y amigos:
Carlos Daniel Gómez González y
Perla Sosa Zaragoza
Por compartir trabajo académico y profesional en la BCCT-UNAM así como los buenos momentos.

Dedicatorias

A mi padres Eloy Ricardo Castro y María de Lourdes Escamilla, que me han enseñado a ser mejor cada día y por estar a mi lado.

A mis hermanos Karina y César, que siempre he contado con su apoyo incondicional y me han dado el ánimo de seguir siempre adelante.

A Saúl Armendáriz Sánchez, un ser humano extraordinario que ha compartido conmigo tiempo de vida así como conocimientos y hemos llevado a cabo proyectos en común. Gracias Saúl por tu dedicación y observación en este trabajo.

A mis sobrinos obedientes, Richi, Aitana, Gloria y Aldair, que este trabajo les motive para que en algunos años lea el suyo. Así que para adelante.

Y a todos mis peludos de cuatro patas ;)

TABLA DE CONTENIDO

INTRODUCCIÓN	3
ÍNDICE DE FIGURAS	9
CAPÍTULO 1	
LOS REPOSITORIOS DE DATOS CIENTÍFICOS	11
1.1. Las iniciativas del Acceso Abierto	13
1.2. Antecedentes de Repositorios	21
1.3. Definición, objetivos, características	24
1.4. Tipología de repositorios	29
1.5. Los datos científicos	31
1.5.1. Antecedentes del Repositorio de Datos	38
1.5.2. Definiciones, objetivos y características del Repositorio de Datos	40
1.5.3. Tipos y alcances	43
1.5.4. Repositorio de datos en Ciencias de la Tierra	48
CAPÍTULO 2	
DESARROLLO DE REPOSITORIOS DE DATOS CIENTÍFICOS	52
2.1. Los datos científicos de investigación	61
2.2. El Programa de Repositorio Institucionales del Consejo Nacional de Ciencia y Tecnología de México (Conacyt).	64
2.3. Los recursos asignados	66
2.4. La selección del software y hardware	68
2.5. El diseño de la estructura técnica usando metadatos:	70
2.5.1. Dublin Core	73
2.5.2. DataCite y OpenAIRE	75
2.5.3. GeoNetwork	83
2.6. La generación de comunidades y subcomunidades	84
2.7. Poblamiento de datos	86
2.7.1. Migración	87
2.8. La estructura de despliegue	88
2.9. DSpace: su estructura	89
2.9.1. El diseño final	92
2.9.2. Diseño web	93

CAPÍTULO 3	
ELABORACIÓN DE UN REPOSITORIO DE DATOS PARA LA INVESTIGACIÓN EN CIENCIAS DE LA TIERRA	95
3.1. Planeación	97
3.2. Contenido	101
3.3. Diseño del DSpace para el Repositorio	105
3.4. Estructura	109
3.5. Comunidades	113
3.6. Metadatos	115
3.7. Software de interface	129
3.8. Despliegue de la información	130
3.9. Diseño web	135
CONCLUSIONES	139
BIBLIOGRAFÍA	144
ANEXOS	150

INTRODUCCIÓN

Los paradigmas de la investigación están en constante evolución y en este siglo la ciencia intensiva de datos y las metodologías asociadas requieren de colecciones de datos variadas, completas, bien documentadas y de gran tamaño para facilitar el análisis y el descubrimiento de patrones y fenómenos tanto en el área de las ciencias, sociales así como de las humanidades.

Para poner a disposición de los usuarios estos datos, es necesario contar con un contexto para entender su origen, organización, sus propósitos, resultados y una serie de cuestiones técnicas que permitirán su acceso. Ya sea a través de Repositorios de datos o de publicaciones relacionadas con ellos que ofrezcan validez y credibilidad a las investigaciones, y puedan ser citadas o por aquellos que reutilicen los datos para generen nuevas investigaciones.

Al hablar del desarrollo de la ciencia en México, y sobre todo al referirnos a las Ciencias de la Tierra –en donde su producción científica nacional es de aproximadamente de 25,500¹ documentos (equivalente al 42% de la producción que se realiza en América Latina), además de su difícil acceso a ella, ya sea a través de plataformas o estructuras, aunándole sí este acceso es por suscripción, registro, pago o que sean financiados con recursos públicos–, el acceso a ella no es tan fácil.

Esta área científica ha crecido rápidamente en la última década, convirtiéndose en materia prima con un parámetro de desarrollo económico entre los países, en donde se invierten grandes cantidades de recursos para su desarrollo por ser prioritaria para la sociedad. Un problema al que se enfrenta nuestro país es a la falta de difusión de información científica y tecnológica que se deriva de una serie de investigaciones que se llevan a cabo tanto por entidades públicas, privadas y de investigación.

Bajo este panorama y debido al gran impacto de los diversos fenómenos naturales que han golpeado a México, así como el aumento de datos para la investigación e información académica, se considera necesario facilitar a la sociedad herramientas que facilitan la consulta de éstos datos de manera libre, oportuna e intuitiva para entender los cambios que sufre el orbe desde el punto de vista de las Ciencias de la Tierra. Temas como el combate al calentamiento global, desarrollo sustentable, cambio climático, variación de oleaje en los mares, movimientos sísmicos, estados geotérmicos en el país, así como de clima espacial, entre otros, son estudiados por los científicos, y que en ocasiones como resultado desarrollan infraestructuras tecnológicas en beneficio de las ciencias de la tierra.

Ante la producción masiva de datos, se requiere facilitar el acceso, el intercambio, almacenamiento, diseminación, así como la difusión de información especializada

¹ Se llevó a cabo una búsqueda de información en el Sistema de Administración del Conocimiento Web of Science sobre la producción científica en Ciencias de la Tierra. Septiembre 30, 2019.

en Ciencias de la Tierra, el Mar y la Atmósfera, mediante herramientas funcionales que permitan a los investigadores y a la sociedad consultar de manera abierta sus líneas de acción y llevar a cabo un mejor estudio de las variantes que se presentan en México y en otras partes del mundo, a fin de contar con información retrospectiva, actual y en tiempo real, producida en diversas entidades especializadas en Ciencias de la Tierra.

Por ello, es importante estructurar una investigación que apoye el desarrollo y la preservación de la ciencia por medio de una serie de procesos y flujos de trabajo de la información (datos), generados por diversos organismos, servicios nacionales, oficiales y de investigación, que abordan temas de prioridad nacional, como los desastres naturales, cambio climático, movimiento sísmicos por ejemplo, los cuales se han incrementado en la los últimos años y cada vez son más cercanos en tiempo.

En ese sentido, el personal académico de la Biblioteca Conjunta de Ciencias de la Tierra (BCCT)² participa de forma activa, en diseñar un Repositorio Institucional de datos, acorde con los Lineamientos Generales de Ciencia Abierta y sea totalmente interoperable con el Repositorio Nacional (RN) del Conacyt, con la finalidad de poner al alcance de especialistas y público en general los datos que se producen en el área de las ciencias de la tierra.

Se considera substancial la creación de los RI de datos para la investigación, por la naturaleza de lo que se produce en las Ciencias de la Tierra. Es importante sumarse a un Proyecto Nacional a fin de contribuir al desarrollo de la ciencia abierta en nuestro país, que permita el libre acceso a la información de estas áreas del conocimiento y facilite el intercambio y la cooperación en proyectos académicos con otras instituciones, así como en el monitoreo e investigación a nivel nacional e internacional de temas prioritarios, incorporando datos, mapas, cartas, sismogramas, mareogramas, fotos, entre otros, mismos que deberían ser consultados en un RI para que la sociedad pueda tener acceso de manera transparente, convirtiéndose en referentes, que van desde el registro de la actividad sísmica, así como de la difusión y divulgación de la información sismológica, hasta la temperatura geotérmica de un pozo, manantial, ojos de agua o, en su caso, el clima espacial que envuelve a nuestra estratósfera.

Si consideramos que México cuenta con 27,188³ investigadores, según el Portal de Datos del Sistema Nacional de Investigadores (SNI), y el total de la producción científica en Ciencias de la Tierra (27.09%) es generada por la UNAM, se

² Biblioteca Conjunta de Ciencias de la Tierra (BCCT), conformada en 1996 con recursos documentales, humanos, económicos y materiales de cinco entidades académicas del área de la Investigación Científica, las cuales son los Institutos de Geología; Geofísica; y Ciencias del Mar y Limnología, así como los Centros de Ciencias de la Atmósfera; y Geociencias (en Juriquilla, Qro.). Cuenta en su contexto con bibliotecas en las Unidades Académicas de Hermosillo, Son., Mazatlán, Sin., El Carmen, Camp., y Puerto Morelos, Q. Roo, y ofrece servicios de información a los académicos que hacen uso de los Buques Oceanográficos Universitarios Justo Sierra y Puma.

³ Conacyt. Sistema Nacional de Investigadores (SNI). Tabla de Investigadores Vigentes 2017 por Grado, Nivel, Género, Adscripción, Entidad Federativa y Área de Conocimiento. Disponible en <https://datos.gob.mx/busca/dataset/sistema-nacional-de-investigadores>. (Consultado: 8-07-2019).

necesitan mecanismos y recursos que puedan apoyar la difusión de toda esta producción científica mediante instrumentos que permitan el manejo, estudio, análisis y uso de los datos, lo que proporcionará un fortalecimiento de información, la cual puede prevenirnos de cualquier eventualidad sobre algún desastre natural.

Contar con un Repositorio Institucional con información especializada en datos continuos servirá de apoyo a la toma de decisiones ambientales y climáticas. Convirtiéndose en un elemento prioritario de México, por su amplio territorio y variedad de áreas geográficas.

Por ello, se pone a disposición de los usuarios elementos clave de consulta libre que funcionan como fuente de información y enlaces con otros repositorios del mundo que ofrezcan una mayor cantidad de datos bajo un concepto de *ciencia abierta*.

Tomando en cuenta el panorama anterior, el presente trabajo partió de las siguientes preguntas:

1. ¿Por qué es importante el desarrollo de RI de datos científicos que apoyen a la investigación de Ciencias de la Tierra?
2. ¿Qué importancia tiene sumarse a un proyecto de alcance nacional e internacional que impacte en la ciencia abierta?
3. Con la gran cantidad de datos que se producen en la investigación de Ciencias de la Tierra, ¿es importante generar mecanismos que apoyen la organización, difusión y preservación de los datos?
4. ¿Es posible lograr una visibilidad de los datos con el desarrollo de RI?
5. ¿Se requiere de personal profesional que se sume y contribuya en el desarrollo, estructuración y toma de decisiones de un RI?
6. ¿La ciencia abierta en México por medio de los RI abrirá un espacio en el uso y manejo de los datos entre las comunidades científicas?

Estas preguntas originaron se planteara el siguiente objetivo general de la investigación:

- Establecer los elementos necesarios que desde el punto de vista bibliotecológico permitan desarrollar un repositorio institucional de datos científicos que apoye a la investigación especializada en Ciencias de la Tierra, con información vigente y bajo una estructura técnica interoperable a nivel mundial.

Asimismo, los objetivos específicos son los siguientes:

- Establecer políticas para crear un mecanismo que gestione los datos para la investigación en Ciencias de la Tierra, según su naturaleza (estaciones, redes, ubicaciones geográficas, espaciales, temperaturas, manifestaciones, etc.), con base en las necesidades actuales de los usuarios de acuerdo con la experiencia que se tiene en la prestación de servicios.

- Diseñar las pautas para organizar y almacenar datos retrospectivos, actuales y en tiempo real, de eficiente localización, especializados en Ciencias de la Tierra.
- Establecer los elementos para llevar a cabo una adecuada preservación digital de los datos para la investigación que se incluyen en un RI.
- Contar con un Repositorio Institucional que abra las puertas para que se desarrollen proyectos científicos en colaboración, ya sea este nacional o internacional, especializado en Ciencias de la Tierra.
- Aumentar la visibilidad de los datos científicos mediante la difusión y distribución por medio de un RI.

Los supuestos que establecemos son los siguientes:

- Con base en la Iniciativa de integración de Acceso Abierto de la información científica y tecnológica en el Capítulo 10 de Ciencia y Tecnología, se requiere de la creación de un Repositorio Institucional de datos que albergue, organice y difunda los datos que se generan en el área de Ciencias de la Tierra e impulse y contribuya en la estrategia nacional para democratizar del acceso a la información.
- Con el aumento de la producción de los datos científicos, deben ser procesados de forma interoperable para que puedan ser localizables y generen nuevo conocimiento bajo un ambiente de Acceso Abierto y de política nacional de información científica con el fin de ir más allá de lo que ofrecen las publicaciones tradicionales.
- En los últimos años, las Ciencias de la Tierra crecen rápidamente, por lo que se requiere de una buena gestión, distribución y difusión de los datos que, por su especialidad, deben ser trabajados en investigaciones científicas bajo las políticas de Acceso Abierto, a fin de potenciarlas y aumentar la visibilidad a nivel nacional e internacional.
- El generar este tipo de plataformas como el Repositorio Institucional de datos, cambiará el hábito del depósito y manejo de éstos entre los investigadores y los especialistas.
- Es un hecho que, que como bibliotecólogos debemos innovarnos con nuevos perfiles profesionales, obligándonos a adentrarnos, en otras disciplinas y será clave para trabajar en la ciencia abierta y lograr el establecimiento de un producto de impacto nacional.
- Finalmente con la conformación de un Repositorio Institucional de Datos a mediano plazo –creado con recursos públicos en México– permitirá su crecimiento constante sin que su contenido se duplique con otras fuentes secundarias de información.

El método que se emplea en el desarrollo de este trabajo académico es dialéctico-analítico, pues, por un lado, implica comprender la importancia de los diversos fenómenos de la naturaleza, la sociedad y del pensamiento, mismos que conllevan las Ciencias de la Tierra bajo sus leyes y acciones. Es importante también llevar a cabo el proceso de identificación de cada una de las partes que caracterizan el

objeto de estudio –en nuestro caso, el desarrollo de un RI de datos–, por lo que se establece la relación causa-efecto entre los elementos (fases) que compone el objeto de la investigación. Partiendo de la recopilación de datos, análisis de información, de normas y estándares nacionales e internacionales para contar con un panorama sobre la situación actual del estado del arte del tema y tener un conocimiento en las bases de una proyección adecuada, se mostrará un caso real que pueda contribuir al desarrollo de otros proyectos académicos, mismo que puede servir de instrumento y guía para otras investigaciones mediante la aplicación de conocimientos teóricos-prácticos del bibliotecario, una visión proceso-administrativa y estructural de la información con la que debe contar un RI, donde las fases del desarrollo son:

- a) Primera: Se revisará y analizará la bibliografía y herramientas de consulta impresas y electrónicas para definir las bases teóricas que conformen los lineamientos del proyecto.
- b) Segunda: Se llevará a cabo la planificación de las políticas y directrices a nivel nacional e internacional sobre ciencia abierta, para establecer los fundamentos del RI.
- c) Tercera: Se presentará la definición del RI, la organización de la información (naturaleza de los datos) y determinación de los metadatos, así como el estudio de diversas licencias y derechos de autor.
- d) La cuarta y última fase: Abarcará la evolución del proyecto, su puesta en marcha y revisión de la interoperabilidad, ingreso de los datos, actualización y diseño web e imagen institucional, lo que hace distintivo a cada Repositorio para su enlace al RN.

Esta investigación cuenta con tres capítulos enfocados en el desarrollo de un Repositorio Institucional de datos científicos, los cuales se describen a continuación:

En el primer capítulo, “Los Repositorios de Datos Científicos”, se centra en los orígenes del Acceso Abierto y se revisan las diversas iniciativas del movimiento, las cuales han tomado fuerza entre las instituciones académicas y científicas, buscando favorecer y promover el acceso libre dando pie al nacimiento de los Repositorios de Literatura. Se presenta un panorama actual del estado que guarda el *Open Access* a nivel mundial. También se ofrece una tipología de los diversos Repositorios Institucionales. Asimismo, se considera el tema de los datos científicos, partiendo de las declaraciones de la *American Geophysical Union* (AGU), se presentan algunas definiciones, objetivos, así como los principios que los gestionan, además de las diversas características, pero sobretodo enfocándonos a los Repositorios de datos en Ciencias de la Tierra.

En el capítulo dos, “Desarrollo de Repositorios de Datos Científicos”, se muestra la apertura de los datos de investigación como los demanda la sociedad y la participación del Gobierno y Financiadores, por lo que se menciona también el Programa Repositorios Institucionales de Consejo Nacional de Ciencia y Tecnología de México. Se aborda el *data sharing*, que es el compartir los datos de

investigación entre los científicos e interesados, con el fin de maximizar esfuerzos y recursos en apoyo al desarrollo de la investigación y promover la innovación y la reutilización de ellos lo que facilitará el trabajo en colaboración entre usuarios, creadores de datos y reutilizadores. En este apartado se muestran normas, técnicas y elementos que describen el contenido, su calidad y condición, con la finalidad de localizar y recuperarlo para, posteriormente, utilizarlo, administrarlo y establecer una comunicación entre las diversas plataformas con las que se requiera interactuar.

En el capítulo tres, “Elaboración de un Repositorio de Datos para la investigación en Ciencias de la Tierra: una propuesta a seguir”, se presenta paso a paso el desarrollo de un Repositorio Institucional de datos para apoyar la investigación en Ciencias de la Tierra con base en la normatividad y lineamientos de una Entidad Federativa mediante el cual se pretende lograr la participación en un Proyecto Nacional que impulse y difunda la ciencia abierta en México. Se muestra un modelo de planeación, así como el desarrollo de actividades técnicas llevadas a cabo a través de diversas estrategias y acciones. Se muestra un caso real mediante el empleo de software libre para facilitar el flujo de trabajo. Se expone el empleo de normas y estándares internacionales, los metadatos, que definen una estructura detallada, posibilitando el intercambio de los datos, así como un correcto diseño web, despliegue y uso de ellos, parte esencial en el momento del diseño y definición del proyecto.

Con la finalidad de mostrar en esta investigación, que existen escasos Repositorios Institucionales de Datos interoperando con el RN, se presenta el resultado de una evaluación de 96 RI considerando características preponderantes que pueden apoyar y orientarnos en el desarrollo y diseño de esta herramienta.

Finalmente, se ofrecen las conclusiones. En los “Anexos” se encuentra información graficada a detalle de la extracción y el análisis de los datos de los RI de Datos y Literatura que están activos y en continuo poblamiento de nuestro país.

Índice de figuras

- Fig. 1.** Modelo de curación del ciclo de vida (DCC Curation Lifecycle Model, 2019, pág. 33).
- Fig. 2.** Modelo de datos FAIR (Hodson, Simon, 2018, pág. 62).
- Fig. 3.** Elementos Dublin Core (pág. 74).
- Fig. 4.** Niveles obligatorios para las propiedades de metadatos DataCite (DataCite Metadata Schema, pág. 77).
- Fig. 5.** Estructura del Repositorio con base en el software libre de DSpace (pág. 86).
- Fig. 6.** Comunidades, colecciones e ítems con permiso (pág. 91).
- Fig. 7.** Tipo de datos que se trabajan en el RI (pág. 98).
- Fig. 8.** Etapas de la planeación para el desarrollo de un Repositorio de Datos (pág. 99).
- Fig. 9.** Modelo de planeación general en el desarrollo de un Repositorio Institucional de Datos (pág. 101).
- Fig. 10.** Licencia Atribución-NoComercial-SinDerivar 4.0 (pág. 103).
- Fig. 11.** Condiciones de la publicación en la revista (pág. 104).
- Fig. 12.** Política de la publicación de los datos (pág. 104).
- Fig. 13.** Política de la publicación de los datos: Metadatos y XML (pág. 105).
- Fig. 14.** Etapas en el diseño de DSpace (pág. 106).
- Fig. 15.** Configuración de DSpace (pantallas de instalación) (pág. 106).
- Fig. 16.** Configuración de hoja de trabajo (pág. 107).
- Fig. 17.** Hoja de trabajo con campos (pág. 107).
- Fig. 18.** Configuración de metadatos (pág. 107).
- Fig. 19.** Visualización de los metadatos (pág. 107).
- Fig. 20.** Prefijo de los metadatos metadataPrefix=oai_datacite (pág. 107)
- Fig. 21.** Roles de reglas entre los administradores, revisores y publicadores (pág. 108).
- Fig. 22.** Creación de Comunidades, según tipología de los datos (pág. 108).
- Fig. 23.** Configuración para visualización campos (pág. 108).
- Fig. 24.** Visualización de las miniaturas de portadas e imágenes (pág. 108).
- Fig. 25.** Protocolo OAI desde interfaz pública (protocolo OAI-PMH del Repositorio Institucional de Datos del Servicio Sismológico Nacional, 2019, pág. 109).
- Fig. 26.** Personalización de interface y complementos de Google Analytics (pág. 109).
- Fig. 27.** Estructura del Repositorio Institucional y su relación con el Repositorio Nacional (pág. 110).

- Fig. 28.** Organigrama de trabajo para el Repositorio de Datos (pág. 110).
- Fig. 29.** Organización de los grupos para desarrollar las actividades (pág. 111).
- Fig. 30.** Roles de los participantes dentro de la plataforma de DSpace (pág. 111).
Presentación de la autorización y revisión del trabajo de los participantes dentro de la Plataforma de DSpace
- Fig. 31.** (Roles de autorización y revisión que se utilizan en el Repositorio de Recursos Geotérmicos para México, pág. 112).
- Fig. 32.** Nivel jerárquico entre la comunidad, subcomunidad, colección y el ítem (pág. 113).
- Fig. 33.** Ejemplo de las comunidades del Repositorio Institucional de Recursos Geotérmicos para México (pág. 115).
- Fig. 34.** Registro de la Comunidad de Guadalajara (Comunidades geográficas) en el Repositorio Institucional de Recursos Geotérmicos para México (pág. 115).
- Fig. 35.** Comunidades del Repositorio Institucional del SSN de México (pág. 115).
- Fig. 36** Registro de campos de metadatos (pág. 128).
- Fig. 37** Para añadir más campos de metadatos (pág. 128).
- Fig. 38** Estructura de software con el enlace del dato (pág. 130).
- Fig. 39** Estructura de despliegue de la información (pág. 131).
- Fig. 40** Presentación del registro del dato en el RI, en una primera fase en el software de DSpace (pág. 132).
- Fig. 41** Estructura de despliegue de la información con un tipo de software para la presentación de los datos (pág. 132).
- Fig. 42** Presentación de otro software para la consulta gráfica del dato (pág. 133).
- Fig. 43** Presentación de la organización de los datos en el RI de Geotermia, en el software de DSpace (pág. 133).
- Fig. 44** Estructura de despliegue de la información en el RI de Geotermia (pág. 133).
- Fig. 45** Presentación del dato, preparado para su reutilización en formato "csv" (pág. 133).
- Fig. 46** Otro ejemplo de RI de datos, presentación del dato (pág. 134).
- Fig. 47** Estructura de despliegue de la información (pág. 134).
- Fig. 48** Presentación del dato (pág. 134).
- Fig. 49** Diseño web: estructura, etiquetado y navegación (pág. 136).

CAPÍTULO UNO

Los Repositorios de Datos Científicos

El aumento de la actividad científica produce cambios profundos a nivel mundial con sofisticados avances en Tecnologías de Información y Comunicación y en el incremento de la inversión, tanto pública como privada. El avance en áreas como las Ciencias de la Tierra, el Mar y la Atmósfera, así como en la Astronomía, la Física o las Ciencias Biomédicas y de la Vida está permitiendo la obtención y la recolección de datos, que, además, crecen de manera exponencial. Como consecuencia, tanto investigadores como organizaciones dedicadas a la investigación se encuentran cada vez con más dificultad para gestionar, analizar e interpretar esta producción de datos en grandes volúmenes.

Podemos decir que la investigación científica en todas las disciplinas está pasando a ser una materia prima y un parámetro de desarrollo económico entre los países, en donde el gobierno invierte importantes cantidades de recursos económicos y humanos para su desarrollo, con el objetivo de poner los resultados al alcance de la sociedad, no importando su nivel académico, siendo principalmente las instituciones de educación superior y los centros de investigación los que deben contar con herramientas de fácil y libre acceso con productos de investigación concentrados en un solo lugar y que puedan ser consultados de forma tanto local como remota por toda la comunidad académica nacional e internacional.

Es por ello que en las últimas tres décadas se ha buscado trabajar en colaboración con grupos de investigación, reestructurando una nueva manera de investigar. En esta época aparecen nuevos servicios en bibliotecas, así como formato y tipos de información –como los set de datos o los llamados “*dataset*”, el software o los resultados de simulaciones, modelos de sistemas, enlaces de objetos con datos, que deben ser almacenados, preservados y difundidos para la generación de nuevo conocimiento–.

La ciencia ha avanzado a pasos agigantados ejecutando nuevas investigaciones cuyos resultados deben ser publicados y ofrecidos a una comunidad científica que pueda apoyar así la multidisciplinariedad y, con ello, la obtención de nuevos datos. Es importante mencionar que si un investigador no tiene la capacidad de acceder a una publicación científica o a un conjunto de datos, puede estar perdiendo información valiosa para su trabajo diario. La publicación y difusión de los resultados de las investigaciones es, por tanto, una necesidad obligada, pero que debe hacerse mediante lineamientos, condiciones y licencias que a continuación se describen.

Es por eso que comenzaremos por centrarnos en los orígenes del acceso abierto, abordando declaraciones e iniciativas con el fin de hacer referencia en el nacimiento de los Repositorios de Literatura que son todos aquellos que albergan en su mayoría documentos científicos y académicos así como diversos recursos. Y como parte de la evolución de ellos y al rápido y cambiante avance de la investigación científica surge la necesidad de desarrollar Repositorios de datos que alojarán y preservarán datos científicos generados durante el proceso de investigación naciendo en las diversas aéreas de la ciencia.

1.1. Las iniciativas del Acceso Abierto

El Acceso Abierto al conocimiento científico empezó a debatirse de manera formal desde la década de los sesenta, cuando el Departamento de Educación y la Biblioteca Nacional de Medicina de los Estados Unidos lanzaron el *Educational Resources Information Center* (ERIC) y con la Biblioteca Nacional de Medicina (NLM) se presenta MEDLARS. Pero los esfuerzos internacionales en beneficio del acceso a la información científica más significativos surgen casi 40 años después, cuando se establece un marco común internacional para el Acceso Abierto.

En los últimos años, el movimiento “*Open Access*” (OA) o *Acceso Abierto* ha tomado mucha fuerza entre las instituciones académicas y científicas. Por lo que este concepto se relaciona directamente a publicaciones de investigación, que busca favorecer y promover el acceso libre y sin restricciones a los trabajos generados por la comunidad científica a nivel mundial.

Estas publicaciones deben cumplir dos condiciones para participar en OA:

1. El autor o autores, así como el titular o titulares que cuentan con los derechos de autor, deben permitir a los usuarios el acceso y autorizar una copia, uso, distribución y transmisión de la obra públicamente, en cualquier medio digital, con una finalidad responsable y sujeto al reconocimiento de autoría.
2. Depositar una versión del documento completo en formato electrónico en un Repositorio que utilice normas y técnicas de los Archivos Abiertos.

Las Declaraciones

Dentro de este contexto, se analizaron y revisaron las diferentes Iniciativas y Declaraciones del Acceso Abierto “*Open Access*” que se han generado en las últimas décadas, presentando y en listando sus acciones y estrategias:

En la Iniciativa de Acceso Abierto de Budapest (2002) (“*Budapest Open Access Initiative*”,⁴ BOAI), en cuanto a su lanzamiento de campaña de lo que hoy conocemos como “*Open Access*” (OA), fue la primera iniciativa en utilizar este concepto. Se proponían estrategias a favor del Acceso Abierto a todas las disciplinas y países, además de ir acompañada de un financiamiento significativo. Las dos principales estrategias por la BOAI fueron a través de repositorios – conocido como Acceso Abierto vía verde– y mediante las revistas con pago denominado Acceso Abierto de la vía dorada.

En ese momento el Acceso Abierto buscaba extenderse aún más renovando su concepto y adaptándose a los cambios en las Tecnologías de la Información. El nuevo reto se convertía en el método de distribución de la producción científica revisada por pares en todas las disciplinas y países. Dentro de la BOAI se hicieron

⁴ BOAI Forum. *Budapest Open Access Initiative* (2002). Disponible en <http://www.soros.org/openaccess/read.shtml/>. (Consultado 20-10-2018).

recomendaciones sobre políticas, en donde cada institución debería tener una política que asegurara las publicaciones y, una vez revisadas por pares, fueran aprobadas y depositadas en el Repositorio de libre acceso.

Se habla también sobre las licencias y reutilización, recomendando el uso de las *Creative Commons* (CC) en sus diferentes equivalencias para la publicación, distribución y uso de los trabajos científico-académicos. Además, cada registro depositado debería ir acompañado de sus metadatos correspondientes y que estuvieran disponibles lo más pronto posible.

En cuanto a la infraestructura y sostenibilidad, dice la iniciativa que cada institución de educación superior debería disponer de un Repositorio y/o participar en un conjunto, manejándolo como un derecho. Y cada uno de ellos estaría obligado a permitir la descarga, uso y citación mediante herramientas que permitieran calcular las métricas alternativas de impacto. Las instituciones dedicadas a la investigación, incluidas las que financian, estarían obligadas a apoyar el desarrollo y mantenimiento de estas herramientas, directorios y recursos esenciales para el progreso y la sostenibilidad del Acceso Abierto.

Finalmente, también consideran la promoción y coordinación con las editoriales, editores, revisores e investigadores para que sean conscientes de los estándares para la publicación en Acceso Abierto, por ejemplo, en materia de licencias, del proceso editorial, de la solicitud de trabajos, de la declaración de la autoría y acerca de cómo gestionar las tasas de publicación.

La *Declaración de Berlín sobre Open Access* (“Acceso Abierto al Conocimiento en Ciencias y Humanidades” en octubre de 2003) establece dos condiciones para el Acceso Abierto:

- 1) El autor(es) y depositario(s) de la propiedad intelectual deben garantizar a todos los usuarios por igual el derecho gratuito, irrevocable y mundial de acceder a un trabajo erudito; lo mismo que licencia para copiarlo, usarlo, distribuirlo, transmitirlo y exhibirlo públicamente, además de distribuir trabajos derivados en cualquier medio digital para cualquier propósito responsable.
- 2) Una versión completa del trabajo y complementarios, en formato electrónico estándar debe depositarse y publicarlo en por lo menos un repositorio que utilice estándares técnicos que definan el Acceso Abierto, que sea apoyado y mantenido por una institución académica, sociedad erudita, agencia gubernamental o una organización establecida, que busque implementar el Acceso Abierto, distribución irrestricta, interoperabilidad y capacidad archivística a largo plazo.⁵

⁵ *La Declaración de Berlín sobre Open Access* (2003) Disponible en http://openaccess.mpg.de/67627/Berlin_sp.pdf (Consultado 20-10-2018).

En la *Declaración de Bethesda*,⁶ en su reunión del 2003 en torno a las publicaciones, su propósito era estimular la discusión dentro de la comunidad de investigación biomédica sobre cómo proceder rápidamente a proporcionar Acceso Abierto a la literatura científica primaria. El documento está dividido en cuatro secciones: la primera es la definición de *trabajo de publicación de Acceso Abierto*, las otras tres refieren a informes que se realizaron bajo tres líneas de acción y grupos de trabajo:

1. *Instituciones y Agencias de Financiamiento*: las cuales patrocinan y fomentan la investigación científica para promover la creación y difusión de nuevas ideas y conocimientos para el beneficio público, en donde reconocen que la publicación de resultados es una parte esencial de la investigación científica y que los costos de publicación son parte de la investigación.
2. *Bibliotecas y Editores*: son a las que más les interesa el Acceso Abierto como un componente esencial de la publicación científica en el futuro y que los trabajos que informan los resultados de la investigación científica deberían ser tan accesibles y de libre uso como sea posible.
3. *Científicos*: aquellos que realizan investigaciones y el interés en garantizar que los resultados de las mismas se difundan de manera inmediata, amplia y eficaz posible. Ofrecen la oportunidad y la obligación de compartir los resultados de la investigación, las ideas y los descubrimientos libremente con la comunidad científica y el público interesado.

En 2003 y 2004 se produjeron otras declaraciones o manifestaciones internacionales de apoyo al movimiento “*Open Access*”, como las que se presentan a continuación:

Dentro de esta política de Acceso Abierto tenemos en 2020 el “*Wellcome Trust sobre Open Access*”⁷, donde se publica que los datos deben ser abiertos y accesibles para todos, dice: “Nuestra política de Acceso Abierto (OA) está cambiando, se aplicará a cualquier artículo que incluya investigación original revisada por pares”. El objetivo de esta política es asegurar que el conocimiento y los descubrimientos sean compartidos y utilizados de una manera que maximiza su beneficio, más si ha sido financiado con fondos públicos, con la finalidad de proporcionar acceso gratuito y en línea a la investigación publicada y maximizar la disponibilidad y uso de las publicaciones y, sobre todo, asegurándose que la investigación financiada se puede desarrollar.

La *Declaración de Valparaíso*⁸ buscó mejorar las posibilidades de las publicaciones científicas de forma electrónica, así como el proceso de

⁶ *Bethesda Statements on Open Access Publishing* (2003). Disponible en <http://www.earlham.edu/~peters/fos/bethesda.htm/> (Consultado 20-10-2018).

⁷ *Wellcome Trust Position Statement in Support of Open and Unrestricted Access to Published Research* (2003). Disponible en http://www.wellcome.ac.uk/doc_WTD002766.html/. (Consultado 20-01-2019).

⁸ Baiget, T. *The Valparaíso Declaration for Improved Scientific Communication in the Electronic Medium* (2004). Disponible en <https://mx2.arl.org/Lists/SPARC-OAForum/Message/519.html/>. (Consultado 11-03-2019).

comunicación mediante el uso tecnologías al enfocarse más a las revistas científicas de mayor visibilidad y accesibilidad. Pero no solo garantiza los contenidos y el formato, sino que fueran estandarizados e indizados de la mejor manera posible en bases de datos e índices, y que los textos completos estuvieran disponibles de forma inmediata en múltiples Repositorios. Además, considera los indicadores bibliométricos y cientiométricos, asegurando su correcta aplicación en el contexto apropiado para evitar errores. Utiliza modelos de software abierto y fuentes de información que debieran fomentarse para brindar igualdad de oportunidades, reflejadas en los costos de publicación como resultado del impulso hacia la edición y publicación electrónicas (dado el hecho de que los costos del proceso de producción están cada vez más a cargo de autores y lectores), lo que debe conducir inexorablemente a sistemas de divulgación científica que sean abiertos y administrados por la propia comunidad científica.

Esta declaración considera parte fundamental a los bibliotecarios y académicos al ser responsables de enseñar a los estudiantes y usuarios en general cómo evaluar la calidad de las fuentes de información que utilizan. Además de que la comunidad científica debe reunirse para analizar, discutir y proponer normas de publicación en el medio electrónico tan pronto como sea posible.

Por otro lado, en 2005, en la *Declaración de Salvador sobre Acceso Abierto*, los participantes del *International Seminar on Open Access* exhortaron a los Gobiernos y a la Comunidad científica a fomentar el acceso abierto en sus políticas de desarrollo científico. En dicha declaración, exigieron que la investigación financiada con fondos públicos esté disponible de forma abierta y considerar el precio de la publicación como parte de costo de la investigación, además de fortalecer las revistas locales de acceso abierto, los repositorios e iniciativas pertinentes, a fin de promover la integración de la información científica de los países en desarrollo en el acervo del conocimiento mundial.⁹

Refiriéndonos a la región de América Latina en 2018 en la *Declaración de Panamá de Ciencia Abierta*, se plantea la necesidad de reconocer a la ciencia abierta como motor de cambio, empoderando a la sociedad a partir de la reivindicación de sus derechos para producir y aprovechar la ciencia, la tecnología y la innovación. Esta declaración está sustentada en los siete principios del *Manifiesto de Ciencia Abierta y Colaborativa hacia una Ciencia Abierta e Inclusiva por el Bienestar Social y Ambiental*, buscando mejorar la comprensión de los beneficios de la ciencia abierta y plantear la necesidad de desarrollar políticas públicas integrales que atiendan los desafíos de la apertura a lo largo de todo el ciclo de investigación científica.¹⁰

⁹ ICML (2005). *Declaración de Salvador sobre Acceso Abierto: la perspectiva del mundo en desarrollo*. Disponible en <http://www.icml9.org/public/documents/pdf/es/Dcl-Salvador-AccesoAbierto-es.pdf> (Consultado 21-03-2020).

¹⁰ *Declaración de Panamá de Ciencia Abierta* (2018). Disponible en https://hiperderecho.org/wp-content/uploads/2018/11/declaracion_panama_ciencia_abierta.pdf (Consultado 21-03-2020)

En cuanto a la *IFLA Statement on Open Access to Scholarly Literature and Research Documentation*¹¹, busca garantizar el acceso a la información a todas las personas, de conformidad con los principios expresados en la *Declaración de Glasgow* en torno a bibliotecas, servicios de información y libertad intelectual. En donde reconoce la elaboración y aplicación de la investigación en todos los campos y que mejorará el progreso, la sustentabilidad y el bienestar humano.

Su posición del Acceso Abierto a la literatura académica y documentación de investigación es vital para la comprensión de nuestro mundo y para la búsqueda de soluciones a los desafíos globales, en especial para reducir la desigualdad de la información. Así, se garantizará la integridad del sistema de comunicación académica, asegurando que toda investigación estará disponible a perpetuidad.

También se consideraron los principios del *Washington D.C. Principles for Free Access to Science*.¹² Su objetivo es bajar los costos de publicación de documentos científicos y permitir su libre consulta entre la comunidad académica del mundo. Mediante esto se han desarrollado diferentes aseveraciones y posiciones que a la fecha no se han logrado concretar como acciones para el libre acceso.

Hablando de las aportaciones de Europa, vemos la *Declaración de la Alhambra*, redactada en el *Seminario para el Acceso Abierto a la información científica*, en la que participan Portugal, España, Francia, Italia, Grecia y Turquía con recomendaciones de políticas y planes de acción para el desarrollo del acceso abierto. Basadas en llevar a cabo políticas para el fomento del Acceso Abierto a la información científica y enfatizar en estas iniciativas de promoción entre investigadores, responsables políticos e involucrados, también refiere al diseño de los modelos editoriales alternativos y sostenibles, asegurando calidad de las publicaciones en Acceso Abierto y cómo fomentar la creación y uso del Repositorio.

En cuanto al panorama actual del estado que guarda el Acceso Abierto, se encuentran inmersos 158¹³ países, revisando estadísticas e informes. Pocas naciones han tenido éxito, por su entorno y los factores que los rodean en la implementación; en otros casos, el avance ha sido bajo o se encuentran en sus primeras etapas de desarrollo.

Según la UNESCO, el trabajo de Acceso Abierto se divide en cuatro regiones, mostrando los avances y el desarrollo obtenido en los últimos años. Estas son:

¹¹ *IFLA Statement on Open Access to Scholarly Literature and Research Documentation* (2004). Disponible en <http://www.ifla.org/V/cdoc/open-access04.html/>. (Consultado 11-03-2019).

¹² *Washington D.C. Principles for Free Access to Science* (2004). Disponible en <http://www.deprinciples.org/statement.pdf/>. (Consultado 11-03-2019).

¹³ *The Global Open Access Portal (GOAP)* UNESCO. Disponible en <http://www.unesco.org/new/en/communication-and-information/portals-and-platforms/goap/>. (Consultado 11-04-2019).

África

Aunque todavía no es un movimiento importante dentro del continente, Mozambique está dando sus primeros pasos, especialmente en la comunidad académica. La idea de tener un Repositorio de Acceso Abierto llegó a la agenda después de un curso de capacitación de tres días sobre Acceso Abierto y Construcción de Repositorios, en las instalaciones de la Universidad Politécnica en julio de 2008. Cada una de las instituciones involucradas –como las bibliotecas universitarias– se sumaron al proyecto de *Repositorio SABRE*,¹⁴ haciendo lo posible por obtener los mejores contenidos. No hay políticas obligatorias nacionales o institucionales, pero existe un incentivo para unirse al movimiento de Acceso Abierto, dirigido al personal académico y de investigación.

Asia

En esta región se experimentó un rápido avance industrial donde los gobiernos invierten importantes recursos, mostrando una influencia significativa en la productividad científica y en los informes, aunque no de forma directa en revistas de Acceso Abierto. Esta región alberga varios repositorios de instituciones y bibliotecas. Por lo regular, los repositorios se basan en documentos como tesis y disertaciones, mientras que otros son más específicos –como el caso de AGRIS, especializado en los temas agrícolas–. Sin embargo, algunos países han mostrado un progreso constante en el impulso de actividades; otros aún no han comenzado con este proceso, como se muestra en los informes por países en Asia Pacífico. Hoy en día, manejan aproximadamente 1,440 revistas OA publicadas en Asia, las cuales están indexadas, principalmente en India, Indonesia y Australia.

Europa y América del Norte

Ambas regiones se comportan de manera variada, desde los países más depositarios de Acceso Abierto a nivel mundial, así como mandatos de financiamiento nacionales y con legislación en América del Norte, hasta aquellos con conectividad limitada a Internet.

Para varios países de Europa, la iniciativa DRIVER –así como la creación de distintas directrices– repercutió en el desarrollo del Acceso Abierto y en el establecimiento de redes de expertos que apoyaran el proyecto, facilitando la implementación de las prácticas. El trabajo se ha llevado a cabo con dos organizaciones: INASP¹⁵, apoyando el desarrollo internacional para mejorar el acceso, la producción, el uso de la información y el conocimiento de la investigación, con la finalidad de que los países cuenten con los requerimientos necesarios para poder resolver sus desafíos de desarrollo; y la EIFL,¹⁶ que

¹⁴ El repositorio de Acceso Abierto de Mozambique fue creado y lanzado conjuntamente en noviembre de 2009 por el Centro de Forma Jurídica y Judicial, la Universidad de Eduardo Mondlane y la Universidad Politécnica. Actualmente posee 2,519 artículos que incluyen tesis y disertaciones, artículos de revistas y documentos de conferencias de investigadores mozambiqueños.

¹⁴ *Research and Knowledge at The Heart of Development (INASP)*. Disponible en <https://www.inasp.info/>. (Consultado 11-04-2019).

¹⁵ *Research and Knowledge at The Heart of Development (INASP)*. Disponible en <https://www.inasp.info/>. (Consultado 11-04-2019).

¹⁶ *EIFL (Electronic Information for Libraries)*. Disponible en <https://www.eifl.net/>. (Consultado 19-02-2019).

trabajan en colaboración con las bibliotecas para permitir el acceso al conocimiento, aprendizaje, la investigación y el desarrollo comunitario sostenible para que todos los individuos tengan acceso al conocimiento y alcanzar su máximo potencial.

América Latina y El Caribe

En cuanto a esta región, las inversiones y los financiamientos para iniciativas de Acceso Abierto e investigación provienen, directa o indirectamente, de fondos públicos y de los proyectos de cooperación internacional. Su estructura de trabajo está basada en redes de información regional que ofrecen Acceso Abierto a sus resultados de investigación, ubicados en Repositorios de temas regionales y que en sus inicios, allá por los ochenta, solo eran registros bibliográficos. Ahora se extienden a textos completos, así como en portales multidisciplinares regionales, revisados por pares, y repositorios digitales, desarrollados desde finales de los noventa.

Como iniciativas regionales de Acceso Abierto, que se ofrecen sin costo para los autores y usuarios finales, a publicaciones académicas y científicas de texto completo, se encuentran los Portales de revistas multidisciplinarios revisadas por pares. Estas plataformas patrocinadas principalmente por fondos públicos, buscando la visibilidad internacional, indicadores bibliométricos y una mejor calidad de los procesos editoriales, tales como DOAJ y SciELO. También hay portales de revistas en Brasil –como SEER, a nivel institucional de la Universidad de Sao Paulo (USP)–. En México, el Portal de revistas académicas y arbitradas de la Universidad Nacional Autónoma de México (UNAM) tiene como objetivo facilitar la búsqueda y consulta de las revistas e incrementar su visibilidad, presencia e impacto. Por otro lado, la Universidad de Chile tiene un significativo número de revistas de Acceso Abierto, por dar algunos ejemplos.

Estas iniciativas de portales de revistas contribuyen a mejorar la metodología y calidad de los procesos editoriales, la visibilidad de las revistas académicas editadas en América Latina y el crecimiento de la adopción del Acceso Abierto gratuito. Y, más recientemente, los Repositorios Institucionales de centros y universidades públicas y privadas. Publicaciones que impulsan y difunden sus contenidos e investigaciones científicas.

En mayo de 2019, se reunieron los representantes de la *African Open Science Platform*, *América, Coalitions*, *OA2020* y *SciELO*, cinco de las principales iniciativas mundiales de Acceso Abierto en la reunión anual de la *Declaración de São Paulo*, con el fin común de hacer que el conocimiento esté disponible y sea accesible en cualquier lugar, pues este ayuda a resolver los desafíos de la humanidad, mediante cinco premisas:

1. Considerar que el conocimiento académico y científico es un bien público mundial, y cuando se genera con fondos públicos es de libre acceso y se convierte en un derecho universal.

2. Proporcionar un acceso universal, irrestricto e inmediato a la información académica, incluyendo su uso y reutilización por humanos y máquinas.
3. Compartir la creencia de que este objetivo común puede lograrse a través de una variedad de enfoques.
4. Buscar puntos de concordancia entre sus enfoques y formas de cooperar para alcanzar el objetivo común.
5. Y buscar un diálogo activo con todas las partes interesadas, incluidos los investigadores, los financiadores de la investigación, las universidades, las bibliotecas, los editores, las sociedades científicas, los gobiernos y los ciudadanos, para tener en cuenta la diversidad de la comunidad académica mundial.

Este movimiento logra mayor disponibilidad y accesibilidad de los resultados de la investigación científica financiada con fondos públicos y amplía las posibilidades de procesos de revisión por pares, acortando tiempos de publicación y aumentando la transparencia en los trabajos científicos, logrando un alto impacto de la investigación.

Las vías del acceso

El modelo de acceso abierto mejora la velocidad, eficiencia y eficacia de la investigación; aumenta la visibilidad, uso e impacto de la investigación; y hace posible que las comunidades de profesionales, expertos y el público interesado se beneficien de la investigación.

Las vías establecidas por el Proyecto *SHERPA/ROMEO* están constituidas en categorías, determinando el estatus de los documentos que se incluyen o integran en el Repositorio de acuerdo con los siguientes colores:

1. Verde: preprint y postprint.
2. Azul: postprint (*ie*, la versión final posterior a la revisión por pares).
3. Amarillo: preprint (*ie*, la versión previa a la revisión por pares).
4. Blanco: el archivo no está formalmente admitido.

Las vías se identifican por colores –blanca, amarilla, azul y verde–, pero solo por los alcances de lo que implica el desarrollo de un RI, consideraremos la Vía dorada (*Gold Open Access*) para la de publicación en revistas; y verde (*Green Open Access*) que refiere al archivo o depósito de las obras en el Repositorio. A continuación se describen:

La Vía Dorada inmediatez y permanencia a los contenidos también conocida como publicación en acceso abierto o modelo de publicación autor-paga.

Brinda al lector el acceso público y gratuito al artículo final (esto es, tal y como se ha publicado), siendo este acceso inmediato y permanente, después de la aceptación del artículo por parte de comité editorial de la revista. En cuanto a las condiciones de utilización de este contenido, están determinadas por la licencia de uso, que puede ser comercial o no comercial.

Es importante considerar que optar por esta vía supone que, una vez publicado el artículo, los usuarios pueden acceder a él en línea de forma inmediata y permanente, pudiendo ser compartido y utilizado por los lectores según las condiciones establecidas por la licencia de uso.

En la Vía Verde hay varias opciones para elegir, conocida como autoarchivo

Proceso donde el autor (que ha publicado su artículo en una revista de suscripción) deposita su artículo, una vez aceptado (postprint), en una web o repositorio de recursos digitales, sin necesidad de pagar (los costes se cubren con las suscripciones que hagan las bibliotecas o los particulares). Estos repositorios recogen la producción de determinadas áreas de conocimiento a nivel internacional y permiten gestionar, preservar y mostrar su producción científica o de datos. Almacenan, conservan y comparten los datos de las investigaciones.

Es importante considerar que para el Acceso Abierto los autores deben cumplir algunas normativas, como pueden ser principios, estrategias y compromisos para poner a disposición las publicaciones o trabajos académicos al alcance de todos. El acceso a este tipo de información indudablemente posee un enorme valor para la generación de nuevo conocimiento, y, a su vez, para el desarrollo de las distintas disciplinas. Debemos estar conscientes de que si se trabaja con financiamiento público, la información e investigaciones que generan las instituciones se debe poner a disposición de la sociedad, siempre y cuando no perjudique el impacto nación y/o mundial, puntos que se explicarán y desarrollarán más adelante.

Determinar cuál es la vía más conveniente depende de varios factores, entre ellos revisar y analizar las políticas por la cuales estamos inmersos en la institución o entidad para la cual laboramos.

1.2. Antecedentes de los Repositorios

Hace poco más de 20 años de la *Propuesta Subversiva*¹⁷ de Stevan Harnad, investigador de Ciencias Cognitivas y de Acceso Abierto en las universidades de Southampton, en Gran Bretaña, y Montreal, Québec en Canadá, donde invitaba a los investigadores a subir a Internet –conocido en aquel tiempo como la World Wide Web (WWW)– el texto completo de sus artículos publicados y ofrecerlos de forma gratuita al público, pero no fructificó como se esperaba. No es sino en 1991 cuando los físicos trabajaron en un proyecto colaborativo llamado *Boletín electrónico sobre Física Teórica de Alta Energía*¹⁸ o también llamado *Repositorio*, en donde Paul Ginsparg albergó los primeros preimpresos de varios investigadores con el deseo de simplificar el intercambio de manuscritos no

¹⁷ *The Subversive Proposal at 20*. Disponible en <https://poynder.blogspot.com/2014/06/the-subversive-proposal-at-20.html?spref=tw>. (Consultado 15-01-2019).

¹⁸ “ArXiv at 20. Paul Ginsparg, founder of the preprint server, reflects on two decades of sharing results rapidly online — and on the future of scholarly communication”. En *Springer Nature Publishing AG*. Disponible en <https://www.nature.com/articles/476145a>. (Consultado 15-01-2019).

publicados, por lo que apostó a compartir y difundir en línea el futuro de la comunicación académica.

Este repositorio automático era un sistema de alerta para preimpresos de Física, que se implementó poco antes del comienzo de la era de la WWW. El plan original era para aproximadamente 100 envíos de artículos en texto completo cada año, cada uno almacenado durante tres meses, hasta que el sistema de distribución de papel existente pudiera ponerse al día. Es importante mencionar que, por demanda popular, nunca se eliminó nada. En pocos años, se convirtió en un recurso web en el sitio <http://arXiv.org>, en donde actualmente contiene cerca de 1,512,136 impresiones electrónicas¹⁹ en los campos de Física, Matemáticas, Informática, Biología Cuantitativa, Finanzas Cuantitativas, Estadística, Ingeniería Eléctrica, Ciencias de Sistemas y Economía.

En 2018, este repositorio recibió cerca de 140,616 nuevas presentaciones, un aumento del 14% con respecto a 2017. Hubo alrededor de 228 millones de descargas en todo el mundo. Actualmente *arXiv* es un recurso global, con casi el 90% de los fondos de apoyo provenientes de diversas instituciones distintas a Cornell; y el 70% del uso institucional proveniente de países distintos a los Estados Unidos.

Otro de los pioneros en el hospedaje del autoarchivo de documentos fue el *Department of Electronics & Computer Science*, de la Universidad de Southampton, con el depósito de su publicación, debido a la difusión y conocimiento del significado de *Open Access*. Este mismo comportamiento se observó en algunas universidades como en la de Australia, en la Universidad de Tasmania y la Universidad Tecnológica de Queensland.²⁰

Para 2007, en Francia los Repositorios adquieren notoriedad por la *l'Agence Nationale de la Recherche*, disponiendo el Acceso Abierto de los resultados de las investigaciones que financiaba. Al año siguiente, el *US National Institutes of Health* y el *Gobierno de la Comunidad Autónoma de Madrid* plantean la misma iniciativa con el fin de dar visibilidad a los resultados de investigaciones financiadas por instituciones del Estado.

Por otro lado, es importante mencionar que surge el proyecto *SHERPA/ROMEO*,²¹ el cual recoge las políticas de las editoriales con respecto al Acceso Abierto. Esta base de datos utiliza un esquema de codificación por colores para clasificar a los editores de acuerdo con su política de autoarchivo, mostrando a los autores si la revista permite el archivo de preimpresión o impresión posterior en sus acuerdos

¹⁹ “ArXiv.org. Open access to 1,555,075 e-prints in the fields of physics, mathematics, computer science, quantitative biology, quantitative finance, statistics, electrical engineering and systems science, and economics”. Disponible en <https://arxiv.org/>. (Consultado 15-01-2019).

²⁰ “Comparison of content policies for institutional repositories in Australia by Arthur Sale”. Disponible en <https://firstmonday.org/article/view/1324/1244>. (Consultado 11-02-2019).

²¹ “Políticas de copyright de las editoriales y autoarchivo”. Disponible en <http://www.sherpa.ac.uk/romeo/index.php?la=es&fIDnum=i&mode=simple>. (Consultado 11-02-2019).

de transferencia de derechos de autor. Se conforma de universidades dirigidas por la investigación con un interés activo en establecer un ejemplo de concepto de *Acceso Abierto* y mostrar los derechos de autor y las políticas de autoarchivo para las revistas académicas. En 2013, al principio de este estudio, eran 1,241 editoriales las que habían definido sus políticas en este sentido. En 2016, el número de editoriales aumentaron a 2,350. El portal se traduce a varios idiomas (español, húngaro, holandés y portugués) y se amplía considerablemente el número de países editores de revistas.

La iniciativa OAI (*Open Archives Initiative*)

Bajo el contexto de la creación y desarrollo de los Repositorios, surge la iniciativa del *Open Archives Initiative* (OAI), que se aprueba en la Convención de Santa Fe (EE.UU.) en octubre de 1999 y desarrolló una alternativa de promover el libre acceso a la información científica de manera ordenada y normalizada, estableciendo como su objetivo el buscar la interoperabilidad entre distintos servidores para el intercambio de datos y ofrecer soluciones y oportunidades mediante procedimientos prácticos y sencillos. Esta interoperabilidad es la condición mediante sistemas heterogéneos que pueden intercambiar procesos o datos. Se basa en tres modelos de trabajo, según Chowdhury, investigador de la National Science Digital Library (NSDL):

1. Compilación, refiriendo a la indización de los sistemas de información.
2. Recolección para el acuerdo, a fin de compartir determinados servicios.
3. Federación mediante directrices para la utilización de estándares y tecnologías comunes.

Según los niveles en los sistemas de información, esta interoperabilidad puede ser:

1. Sintáctica: para la representación compatible a través de lenguajes de marcado (XML) y esquemas de metadatos.
2. Semántica: para el intercambio de información, significado común del contenido de los elementos de metadatos.
3. Estructural: con la finalidad de comunicarse e interactuar en contextos heterogéneos.

En cuanto a la esencia, la visión de OAI se divide en dos sectores:

- Proveedores de datos para que los repositorios sean accesibles en la red con los recursos y sus metadatos.
- Proveedores de servicios para cosechar²² metadatos y ofrecer servicios de valor añadido.

²² Definición de *Cosecha*, es el proceso que confirma que las estructuras técnicas están normalizados bajo un estándar y son interoperables entre los Repositorios. Su función principal es mover metadatos de un servidor local a un agregador con la finalidad de reunirlos con otros metadatos y proporcionar un servicio de valor agregado. Es importante mencionar que se requiere de un acuerdo entre el proveedor de metadatos y el cosechador.

Técnicamente, todo el funcionamiento del sistema se basa en HTTP, XML y los metadatos Dublin Core sin cualificar (ISO-15836).

Por otro lado, la creación de la interface OAI-PMH (*Open Archives Initiative*) – también conocida como *Protocol for Metadata Harvesting*– permite que los contenidos estén disponibles (primera versión en 2001), por lo que requiere de una normalización y adopción de estándares, que pueden estar basados en organismos como W3C (World Wide Web Consortium), la ISO (International Organization for Standardization), el OASIS (Advancing Open Standards for the Information Society) y el Standards at the Library of Congress.

1.3. Definición, objetivos, características

En este punto se presentan varias definiciones de la literatura, desde una postura general y amplia con el fin de llegar a un concepto particular, buscando se entienda la esencia de la definición de un *Repositorio*, con base en nuestra experiencia y práctica en el desarrollado de Repositorios Instituciones (RI) en México, en particular en la UNAM. Para ello se presenta a continuación el siguiente contexto teórico:

Definiciones

El origen de la palabra *Repositorio* deriva del latín *repositorium*, que significa “armario, alacena”. El *Diccionario de la Real Academia de la Lengua Española*²³ (2001) la define como “Lugar donde se guarda algo”. Y *Wikipedia* (2019)²⁴ explica:

[...] un repositorio es un espacio centralizado donde se almacena, organiza, mantiene y difunde información digital, habitualmente archivos informáticos, que pueden contener trabajos científicos, conjuntos de datos o software.

Esta última acepción relaciona un contexto tecnológico, lenguaje de Tecnologías de la Información y de la Comunicación (TICs), relacionándola con espacios, depósitos de información digital, formatos de diferente naturaleza preparados para ser distribuidos mediante la Web.

Esta concepción, relacionada con el desarrollo de un Repositorio Institucional (RI), emerge en 2002 como una nueva estrategia que permitió a las Instituciones de Educación Superior (IES) asumir el papel de editoras, modernizando los procesos de publicación y divulgación de la producción académica en contenido digital.²⁵

²³ Diccionario de la Lengua Española (2001). Disponible en <https://dle.rae.es/> (Consultado 23-05-2019).

²⁴ Wikipedia. La enciclopedia libre. Disponible en <http://es.wikipedia.org/wiki/Repositorio>. (Consultado 19-042019).

²⁵ Lynch, C. A. (2003), “Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age”, ARL Bimonthly Report 226 (Febrero de 2003), 1-7. (Consultado 18-01-2019).

En ese mismo año, Drake comenta:

[...] el desarrollo de Repositorios Institucionales nace como una estrategia nueva que permite a las universidades aplicar un control sistemático que acelere los cambios, publicando contenido académico a través de medios digitales.²⁶

Bajo este panorama, los RI representan una importante fuente de contenido académico y científico, que materializa las expectativas de acceso libre a la información científica. Los artículos publicados en revistas pueden además ser cargados en el depósito institucional y/o en otros repositorios temáticos. Para ello, las revistas deben seguir una política denominada *Vía verde*.²⁷

Estos conceptos permiten afirmar que los RI ofrecen un conjunto de servicios en línea que facilitan localizar, almacenar, ordenar, preservar, conservar, redistribuir y difundir objetos documentales en diversos formatos para hacerlos accesibles libremente a la sociedad especializada como al público en general.

Raym Crow aporta una definición contemporánea (2002):²⁸

[...] un repositorio institucional, en términos generales, puede ser cualquier colección de material digital alojado, bajo propiedad, o control, o difundida, por un colegio o universidad, independientemente del propósito dado que pueden asumir muchas formas y servir una variedad de propósitos.

En este caso, la definición se centra en un tipo de Repositorio Institucional en particular: un archivo digital del producto intelectual creado por una unidad académica, personal de investigación, estudiantes de una institución y accesible a los usuarios finales, tanto dentro como fuera de la misma institución, con pocos o ninguna barrera de acceso.

Veamos ahora cómo se definen el Repositorio Institucional (RI). Según Clifford Lynch²⁹ (2003):

[...] un RI es un conjunto de servicios que ofrece a los miembros de la comunidad para la dirección y distribución de materiales digitales creados por la institución y los miembros de esa comunidad.

Añade un compromiso para la administración de estos materiales digitales, incluyendo la preservación a largo plazo cuando sea necesaria, así como la organización y acceso o distribución.

²⁶ Drake, M. (2004). "Institutional Repositories: Hidden Treasures". Disponible en <http://www.infoday.com/searcher/may04/drake.shtml>. (Consultado 11-02-2019).

²⁷ En la vía verde, el autor es el responsable de poner a disposición los artículos en un Repositorio Institucional de forma voluntaria. Esta vía pone enteramente en las manos e intereses de los proveedores de la propia investigación y la comunidad mundial de investigación la responsabilidad de la publicación. (Consultado 18-02-2019).

²⁸ "The case for institucionales repositories". Disponible en <http://www.sparc.arl.org/resources/papers-guides/the-case-for-institutional-repositories>. (Consultado 23-02-2019).

²⁹ Lynch, C. A. (2003). Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. Disponible en <http://www.arl.org/newsltr/226/ir.html>. (Consultado 23-03-2019).

El repositorio institucional contempla, esencialmente, una visión de “Acceso universal al conocimiento registrado en formato digital sin límites de tiempo ni espacio”. Volder³⁰ nos dice:

[...] un repositorio institucional reúne la producción científica y/o académica de los miembros de una o varias instituciones, resultado de su actividad docente e investigadora, almacenando, preservando, divulgando y dando Acceso Abierto a los recursos depositados en ellos.

La *SPARC (Scholarly Publishing and Academic Resources Coalition)* menciona que:

Los repositorios adquieren la denominación de *institucionales* cuando reúnen las características puntualizadas, es decir, cuando pertenecen a una institución son de ámbito académico, son acumulativos y perpetuos y son abiertos e interactivos³¹.

Y los tipifica de la siguiente manera:

- Pertenecientes a una institución.
- De un ámbito académico.
- Acumulables y a perpetuidad.
- Abiertos e interactivos.
- Cumplen con el Protocolo OAI-PMH, permitiendo el Acceso Abierto a la documentación académica.

El Informe APEI sobre Acceso Abierto (2008)³² define un Repositorio Institucional (RI) como “una colección de objetos digitales basada en la web, de material académico producido por los miembros de una institución (o varias), con una política definida”.

Actualmente, contamos con la definición de la *American Library Association (ALA)* de los Estados Unidos: “Repositorio es un lugar para objetos digitales creados a partir de la investigación de una institución en específico, que tiene por objetivo acceder y preservar dichos objetos digitales”.

Por lo que en el sentido más simple podemos decir que el término de *RI* es un archivo electrónico de la producción científica de una institución, almacenada en un formato digital, en el que se permite la búsqueda y la recuperación para su posterior uso.

³⁰ “Información, cultura y Sociedad”. Disponible en <http://revistascientificas.filo.uba.ar/index.php/ICS/article/view/833>. (Consultado 30-03-2019).

³¹ Crow, Raym (2002). “The Case for Institutional Repositories: A SPARC Position Paper”, Washington, DC. Scholarly Publishing & Academic Resources Coalition. Disponible en http://www.arl.org/sparc/IR/IR_Final_Release_102.pdf. (Consultado 29-03-2019).

³² Alonso, Julio; Subirats, Imma; Martínez Conde, Ma. Luisa (2008). *Informe APEI sobre Acceso Abierto*. Gijón: Asociación Profesional de Especialistas de la Información. Disponible en <http://www.apei.es/informes/InformeAPEI-Accesoabierto.pdf>. (Consultado 29-03-2019).

Según Villadónica Gómez,³³ los Repositorios también son tratados como:

Organismos públicos y privados, principalmente las universidades, trabajan en la creación, el desarrollo, la alimentación y el mantenimiento de Repositorios Digitales Institucionales, reconociendo en ellos el nuevo paradigma de la comunicación científica, donde confluyen los cauces oficiales y habituales de la comunicación, sumados a las políticas de Acceso Abierto (AA).

Es por eso que los RI están en constante desarrollo dentro del área académica y científica. Los documentos albergados, particularmente los estándares, requieren de mayor flexibilidad por parte de las plataformas de software. Es necesario establecer parámetros en torno a los tipos de contenido que se alojarán en el Repositorio y esto dependerá en gran medida del modelo que estemos interesados en generar.

Hoy en día, las universidades o institutos de investigación son los que, en general, gestionan este tipo de Repositorios, constituyendo una herramienta clave de sus políticas científicas y académicas, además de una pieza de apoyo para la enseñanza y la investigación, por lo que no es correcto llamarle *base de datos*, pues debe de cumplir con una serie de pautas, políticas y lineamientos que más adelante se explicarán.

Desde un punto de vista más conceptual, un RI se convierte en un sistema de gestión de contenidos que ofrece a la comunidad académica un conjunto de servicios para la producción académica.

Finalmente, concluyo que el Repositorio Institucional de Literatura se convierte en una solución de comunicación en línea con el fin de desarrollarse en tres actividades: 1) recopilar, 2) preservar y 3) difundir las publicaciones o ciencias de los colegios, universidades e instituciones de investigación. Y desde hace ya casi veinte años han establecido varias plataformas de Repositorio, cada una con su propio conjunto de beneficios y criterios técnicos. Haciéndolos diferentes de los Repositorios Institucionales de Datos tendiendo como objetivo almacenar y preservar datos científicos que son generados durante el proceso de investigación y que también pueden ser creados por un evento o impacto en la tierra o en su caso recolectado mediante un métodos científico ya sea en tiempo real o mediante un modelo de recolección de datos.

Objetivos

Según los propósitos que se deseen alcanzar en los objetivos y según el tipo de Repositorio que se estudie, analice o desarrolle, las respuestas pueden variar con diversos fines en este proceso; sin embargo, nos basamos en tres puntos primordiales:

³³ Evans & Sutherland Computer Corp. Disponible en <https://es.com/document/388537042/Repositorios>. (Consultado 29-03-2019).

1. Difusión: la cual busca favorecer la difusión de la ciencia y el acceso a la información y/o datos que se crean en una Entidad.
2. Visibilidad: maximiza la visibilidad de la institución mediante su producción científica a nivel mundial.
3. Preservación: con el fin de poner a disposición de la sociedad lo que se produce con fondos públicos por la institución, de manera permanente.

Características

Las cualidades más preponderantes que permite identificar y distinguir entre los diferentes Repositorios son:

- El libre acceso a la información, según el formato que se determine publicar.
- Depósito de autoarchivo, ya sea por el propio autor, creador y/o personal académico autorizado para el desarrollo de este proceso.
- Imprescindible la interoperabilidad, que debe tener la capacidad para funcionar con otros sistemas y/o plataformas que existen en las infraestructuras tecnológicas que se desarrollen bajo el Protocolo OAI-PMH.
- Preservación digital a periodos de largo plazo que garanticen la permanencia y seguridad de la información, manteniendo fiabilidad y legibilidad.

Estas herramientas de administración de contenidos almacenados se describen y apoyan en la preservación de objetos digitales de diferente naturaleza, como pueden ser textos, imágenes, sonidos, datos, etc., con el fin de llevar a cabo su difusión y visibilidad.

Características	Descripción
Estructura de datos	– Material digital que constituye el recurso documento – Información sobre los datos metadatos – Identificador único y persistente del objeto digital
Origen	– Creados en forma digital (born digital) – Digitalizados
Número de documentos	– Sencillos: un solo documento – Compuestos: paquetes de documentos
Almacenamiento	– Local (servidor propio) – Externo (nube)
Acceso	– Público – Forma restringida – Contenidos descritos a través de estándares de metadatos para que puedan ser reutilizados por otros sistemas
Formato	– Texto, imagen, sonido, multimedia, datos
Usos	– Establecer una comunicación y colaboración académica – Publicación digital de los resultados académicos y de investigación

Cuadro personal elaborado con base en la presentación realizada por Ana Poveda de la UC3M (septiembre, 2014).

1.4. Tipología de repositorios

En este contexto de la ciencia abierta, y con base en el desarrollo que tiene esta nueva infraestructura de servicios que ofrecen los Repositorios, existen diferentes tipos de herramientas por la cuales podemos recuperar información y datos científicos que se encuentran disponibles en su mayoría en Acceso Abierto o, en algunos casos, con restricciones por características de la publicación. Se revisó diversa documentación que hablaba sobre tipos de Repositorios, entre ella un documento de Días, Armas, Rodríguez y Camacho,³⁴ pero se complementó con más información y se consideraron los siguientes tipos y ejemplos para una mejor explicación:

- *Repositorios digitales temáticos o disciplinarios*

Se caracterizan por ser creados por grupos de investigadores y gestores de información, en un ámbito específico del conocimiento, con el fin de mejorar el sistema de comunicación entre la comunidad científica y se logró reunir, preservar y dar acceso a los contenidos de un área temática en específico. Algunos ejemplos de ellos son:

- ArXiv: contiene información sobre campos de Física, Matemáticas, Informática, Biología Cuantitativa, Finanzas Cuantitativas, Estadística, Ingeniería Eléctrica y Ciencias de Sistemas y Economía. Disponible en <https://arxiv.org/>.
- E-LIS: sobre temas de Acceso Abierto, especializado en Ciencias de la Información, Bibliotecología y disciplinas relacionadas. Disponible en <http://eprints.rclis.org/>.
- CogPrints: refiere al área de Psicología, Neurociencia y Lingüística, entre otras más. Disponible en <http://cogprints.org/>.

- *Repositorios Institucionales (RI)*

Elaborados por las organizaciones académicas, de investigación, gubernamentales, etc., con el firme objetivo de almacenar, difundir, preservar y facilitar el acceso a la producción científica que generan sus miembros. Ejemplos:

- Repositorio Institucional CICY: información sobre investigación, docencia e innovación que se llevan a cabo en el Centro de Investigación Científica de Yucatán, A.C. Disponible en <https://cicy.repositorioinstitucional.mx/jspui/>.
- Repositorio Institucional de la UAEM, contiene información sobre la producción académica, científica, tecnológica, de innovación y cultural que genera la Universidad Autónoma del Estado de México. Disponible en <http://ri.uaemex.mx/>.
- Repositorio Institucional SISBIUQROO: cuenta con acceso a la producción científica de su comunidad universitaria. Disponible en <http://rasisbi.uqroo.mx/>.

³⁴ Díaz Pérez, M., Armas Peña, D., Rodríguez Font, y Camacho Díaz, A. F. (2015). Política informacional para Repositorios institucionales en universidades. Revista Atlante: Cuadernos de Educación y Desarrollo. En <http://www.eumed.net/rev/atlante/2015/10/repositorios.html> . (Consultado: 15-05-2019).

- *Repositorios huérfanos* (“homeless”)

Son utilizados para depositar y archivar trabajos de autores que no tienen posibilidad de almacenar su producción científica en un repositorio personal o institucional propio. Ejemplo de ellos son:

- Repositorios de Acceso Abierto de Portugal, su objetivo principal es llevar a cabo la recopilación, agregar e indexar contenidos científicos de Acceso Abierto. Este proyecto es una iniciativa de la *UMIC Knowledge Society Agency*, desarrollada por FCCN Fundação para a Computação Científica Nacional, en colaboración técnica y científica de la Universidad de Minho. Disponible en <https://www.rcaap.pt/>.

- *Recolectores de Repositorios*

Estas plataformas se basan en una estructura informática que permiten una comunicación del protocolo OAI-PMH, lo que facilita la recolección de la información, facilitando así la interoperabilidad entre ellos y el acceso a las publicaciones mediante una plataforma común. Estos son:

- SIUBDU2/Argentina. Disponible en <http://cosechador.siu.edu.ar/bdu3/Browse>.
- Recolecta/España. Disponible en <https://recolecta.fecyt.es/>.
- Repositorio Nacional/México. Disponible en <https://www.repositorionacionalcti.mx/>.

- *Repositorios de datos de Investigación (RDR, Research Data Repositories)*

Almacenan y preservan datos científicos generados durante el proceso de investigación, en tiempo real o mediante un modelo de recolección de datos y pueden estar vinculados a una publicación. Pueden existir como repositorios independientes o integrados a repositorios institucionales. Algunos ejemplos representativos son:

- GenBank: esta infraestructura está diseñada para proporcionar a la comunidad científica el acceso a datos sobre la secuencia genética NIH. Disponible en <https://www.ncbi.nlm.nih.gov/genbank/>.
- Addgene: es un Repositorio sin fines de lucro que apoya a los científicos a compartir datos sobre plásmidos, que se utilizan comúnmente en el área de las Ciencias de la Vida. Disponible en <https://www.addgene.org/>.
- Repositorio de datos del Servicio Sismológico Nacional: contiene datos e información en tiempo real sobre los mapas de sismicidad y estaciones sísmicas y eventos paralelos, así como de los movimientos sísmicos de toda la República Mexicana. Siendo uno de los principales a nivel nacional y que apoya de forma directa la investigación. Está dirigido también a la población en general. Disponible en <http://rigeofisica.ssn.unam.mx/jspui/>.

- *Repositorio de Datos para la Investigación (DRR, Data Repositories for Research)*.

Tienen las condiciones y los datos, cuentan con un formato para ser reutilizados con diferentes fines y propósitos docentes y científicos. Están establecidos para generar nuevo conocimiento, descubrir otras líneas de investigación, trabajar bajo una dinámica diferente y relacionar a grupos multidisciplinares de especialistas de un área científica determinada:

- Repositorio de Recursos geotérmicos para México: contiene datos puros sobre la recopilación de descargas de agua subterránea (pozos, manantiales, norias, etc.), de interés para la geotermia en México. Disponible en <http://rigeotermia.geofisica.unam.mx/jspui/>

Por otro lado, Gregoris Mentzas, Dimitris Apostolou, Ronald Young y Andreas Abecker, en su artículo “Knowledge Networking: a Holistic Solution for Leveraging Corporate Knowledge”,³⁵ publicado en el *Journal of Knowledge Management*, establecen, aunque no es el motivo principal de su texto, una clasificación de los repositorios de información que se pueden encontrar dentro de una corporación. Los divide en tres categorías:

1. *Repositorio de Conocimiento Externo*: lo refieren al almacenamiento de la información generada fuera de la corporación, pero relevante para los procesos de producción de la misma. Desde documentación que se encuentra en la Red hasta información generada por los clientes de la empresa y a la que se tenga acceso.
2. *Repositorios Internos Estructurados*: enfocados a la información generada en la empresa que se ha protocolizado y ha dado lugar a estándares de trabajo y definición de procesos, como flujogramas de trabajo, libros de estilo, planes de calidad, protocolos de actuación, etcétera.
3. *Repositorios Internos Informales*: estos contienen información interna que debe ser tratada e interpretada, obligando al usuario a adaptarla según el contexto. Es decir, de conocimiento que no ha devenido en un estándar, algoritmo o protocolo de trabajo.

1.5. Los datos científicos

Partimos de que los *datos* son un término que evoluciona de forma rápida y en los últimos años han crecido de manera exponencial, considerando que pueden ser cualquier información que refiere a un hecho que debe ser analizado. El *Merriam-Webster's* lo define como la “Información objetiva [como mediciones o estadísticas] utilizada como base para el razonamiento, la discusión o el cálculo”.³⁶

³⁵ “Knowledge networking: a holistic solution for leveraging corporate knowledge”. *Journal of Knowledge Management*. Disponible en <https://www.emeraldinsight.com/doi/full/10.1108/13673270110384446>. (Consultado 29-05-2019).

³⁶ Webster, N. *Merriam-Webster Learner's Dictionary*, “Data” Disponible en <http://www.merriam-webster.com/dictionary/data>. (Consultado 29-03-2019).

Con el desarrollo de iniciativas internacionales de investigación a gran escala, el apoyo a los datos tomó una variedad de formas. Alentados por la planificación posterior a la Segunda Guerra Mundial, se establecieron una serie de instalaciones de datos; por ejemplo, los Centros de Datos Mundiales y la Federación de Servicios de Análisis de Datos Astronómicos y Geofísicos evolucionaron, comenzando con el Año Geofísico Internacional (por sus siglas en inglés International Geophysical Year (IGY)) en 1957-1958, con un enfoque en la ciencia internacional. Desde el IGY, surgió una visión revolucionaria en el área científica de las Ciencias de la Tierra, centrando la atención de los geocientíficos en conjunto en métodos científicos, mediciones y datos. El Consejo Internacional de Sindicatos Científicos (ahora Consejo Internacional para la Ciencia) estableció un sistema de Centros de Datos Mundiales para servir al IGY y desarrolló planes de gestión de datos para cada disciplina científica del IGY.

Estos Centros de Datos Mundiales se centraron en replicar los datos en los centros y compartir datos en todo el mundo. En sus inicios contenía datos para IGY en libros, informes, pero más adelante se desarrollaron otras iniciativas. Los primeros esfuerzos de datos crecieron para incluir cintas magnéticas y tarjetas perforadas en lugares designados. Actualmente, la gestión en los centros de datos ha crecido para incluir datos digitales y muestras físicas, así como para dar cabida a muchas partes interesadas y audiencias.

En el área de las Ciencias de la Tierra de los Centros de Datos Nacionales se encuentran el de Datos Climáticos y el de Datos Oceanográficos, que se crearon para respaldar la gestión de datos desde plataformas con grandes flujos de datos, como los satélites. Y desarrollaron recursos nacionales mediante infraestructuras tecnológicas para los investigadores. Estos centros se convirtieron en depósitos de datos junto con otras instituciones de preservación, como archivos con colecciones de fotos y manuscritos, museos con objetos físicos y bibliotecas con libros y revistas.

Actualmente, hay debates científicos sobre la producción de grandes datos y de infraestructuras que se centran en investigaciones, reavivando el interés social y cultural, transformándose en nuevas líneas de investigación y de conocimiento.

La investigación científica está obligando a llevar a cabo un análisis y/o procesamiento e interpretación de datos mediante diversas formas de recopilación, tipo de variable o necesidad de la precisión, según el propósito del estudio o análisis. Estas actividades presentan desafíos para dos grupos de investigadores –quienes usan y producen datos; y para los profesionales de la información, que deben buscar mecanismos para la organización, administración y preservación de los datos–.

En otro contexto de la investigación, los *datos* tienen diferentes concepciones y pueden no relacionarse con un objeto, describirlo o referirlo a una información específica. El *Engineering and Physical Sciences Research Council* (EPSRC) señala:

[...] los datos de investigación son aquellos datos que se definen como material fáctico registrado, comúnmente retenido y aceptado en la comunidad científica, según sea necesario para validar los resultados de la investigación; aunque la mayoría de estos datos se crean en formato digital, todos los datos de investigación se incluyen independientemente del formato en el que se crean.³⁷

Por lo que algunos de los beneficios que implican dicho proceso son:

- Aumentar la visibilidad de la investigación y generar citas, lo que conlleva al crecimiento de la reputación científica de los investigadores individuales, sus equipos de investigación y sus instituciones.
- Protegerse contra el uso de datos erróneos al permitir que los resultados publicados se verifiquen, refuten o refinan de forma independiente, mejorando así la calidad general de la investigación y fomentando la diversidad de análisis y opiniones, y ayudando a resolver disputas científicas.
- Estimular nuevos enfoques para la recolección de datos y métodos de análisis.
- Aumentar la conciencia de la investigación en áreas relacionadas que conducen a más oportunidades de colaboración.
- Facilitar la educación de nuevos investigadores y del público en general.

El modelo de Curación del Ciclo de Vida del Centro de Curación Digital (DCC) señala: “Los datos, cualquier información en formato digital binario, se encuentran en el Centro del Ciclo de Vida de Curación”.³⁸



Figura 1. Modelo de Curación del Ciclo de Vida³⁹

³⁷ Engineering and Physical Sciences Research Council (EPSRC). Disponible en <https://epsrc.ukri.org/about/standards/researchdata/scope/>. (Consultado 18-05-2019).

³⁸ DCC Curation Lifecycle Model. Disponible en <http://www.dcc.ac.uk/resources/curation-lifecycle-model>. (Consultado 18-05-2019).

³⁹ Idem.

Los datos de investigación pueden ser experimentales, observacionales, operativos o pueden ser el resultado de un tercero. Pero además también hay datos del sector público. Y en el caso de las Ciencias de la Tierra su nacimiento es constante e infinito por su forma de obtención, ya sea en tiempo real, por interpretación, análisis, seguimiento, procesados y reutilizados.

Asimismo, se le ha denominado a los *datos abiertos* como “conjuntos de datos cuyos metadatos son obligatoriamente públicos.”⁴⁰ Es decir, cualquier individuo de la sociedad puede consultarlos, usarlos y/o reutilizarlos, pero están en relación con una investigación, artículo y/o objeto que tiene un fin establecido. Al hablar de datos abiertos hay que hacer además la diferencia entre los *datos gubernamentales* y los *científicos*.⁴¹ Los primeros son los que genera y recopila el gobierno bajo transparencia, participación social y disposición de público para que sean consultados y, en algunos casos, reutilizados para impulsar nuevas iniciativas y usos comerciales. Los segundos tienen también un sesgo gubernamental debido a que se obtienen mediante financiamiento público, pero con la diferencia de que se involucran métodos, procesos y sistemas para extraer conocimiento y comprensión del entorno en diversas áreas, ya sea estructurados o no estructurados, lo que implica un análisis y la obtención de resultados de nuevo conocimiento o líneas de investigación.

Mostramos algunos ejemplos basándonos en información que actualmente está hospedada en portales como *Datos Abiertos de México*,⁴² *Datos Abiertos UNAM*,⁴³ *Datos Abiertos de la Ciudad de México*,⁴⁴ *Datos Abiertos 100 México-Open Data 500*,⁴⁵ hay varios tipos y para diversos fines, o según su existencia, que no son obligatoriamente públicos y/o se encuentran restringidos; por ejemplo, en el área de Química, las patentes, etcétera.

Es importante aclarar que a veces los datos no solo se asocian al área de las Ciencias Duras, sino también se encuentran en la parte social o en cualquier otra disciplina y de muchas formas. Puede haber datos obtenidos por métodos de instrumentación o por modelos definidos, simulación o visualización. En el área de humanidades pueden ser incluidos en los textos, en la reconstrucción de hechos, en encuestas por datos de observación o interacción con otros individuos.

⁴⁰ Kaye, John. “Storing and sharing research data after the ‘Space Race’”. Disponible en <https://www.jisc.ac.uk/blog/storing-and-sharing-research-data-after-the-space-race-25-feb-2015>. (Consultado 18-05-2019).

⁴¹ Según Hernández-Pérez y García-Moreno. Disponible en TFG_InfyDoc_GutierrezGallego_Manuel_SI_85_2014-2015. (Consultado: 20-04-19).

⁴² México. Gobierno de México. *Datos Abiertos de México*. Disponible en <https://datos.gob.mx/>. (Consultado: 20-04-19).

⁴³ Universidad Nacional Autónoma de México. *Datos Abiertos UNAM: colecciones universitarias*. Disponible en <https://datosabiertos.unam.mx/>. (Consultado: 20-04-19).

⁴⁴ México. Gobierno de la Ciudad de México. *Datos Abiertos de la Ciudad de México*. Disponible en <https://datos.cdmx.gob.mx/>. (Consultado: 20-04-19).

⁴⁵ Esta iniciativa es el primer estudio de las compañías y ONG’s mexicanas que utilizan datos gubernamentales abiertos para crecer sus negocios, desarrollar nuevos productos y servicios o impulsar el valor social. El Gobierno de la República Mexicana colabora con The GovLab, de la Universidad de Nueva York, para realizar *Datos Abiertos 100 México*. Disponible en <http://www.opendata500.com/mx/>. (Consultado: 20-04-2019).

Dentro de esta investigación, es importante resaltar que puede haber dos tipos de datos:

1. Que por su naturaleza, son aquellos resultados de las investigaciones científicas generados durante el proceso.
2. Y los generados en tiempo real, que pueden ser compartidos y reutilizados por la comunidad científica y la sociedad para generar nuevo conocimiento. Es decir, estos últimos pasarán previamente por diferentes procesos según su utilidad o línea de investigación que está a punto de iniciar, por lo que pueden pasar a ser recolectados, analizados, interpretados y tratados.

Algunos ejemplos pueden corresponder a:

- Peligros naturales, por ser datos creados al momento. Esta asimilación de los datos en tiempo real se está volviendo cada vez más importante para monitorear y predecir cambios en las Ciencias de la Tierra, el Mar y la Atmósfera, así como en otras disciplinas que se encuentran en constante observación de un fenómeno o evento infinito.
- Recolecciones de datos del clima espacial.
- Movimientos sísmicos que tiene la Tierra.
- Datos sobre el manantial o pozo según el ambiente climático.
- La contaminación ambiental que oscila en diversas zonas y niveles de la Tierra.
- Los niveles geotérmicos de la Tierra
- La radiación solar, entre otros.

Es importante considerar que son variados y tienen una gran dependencia tanto de la disciplina del conocimiento a la que pertenecen o en donde se generan, según el ciclo de vida que se le puede dar en una investigación. Esto no resulta tan simple debido a su carácter heterogéneo, a que pertenecen a una dependencia que está sujeta a una variedad de propósitos y procesos de investigación. Por lo que nos damos a la tarea de revisar otras declaraciones que han sido elaboradas y establecidas por diversas organizaciones que están relacionadas en las Ciencias de la Tierra y que tienen una existencia aproximada de 100 años. Estas se describen en el siguiente punto.

La Declaración de la Unión Geofísica Americana (AGU)

Enfocándonos en datos de las Ciencias de la Tierra y el Espacio, de acuerdo con la *American Geophysical Union (AGU)*,⁴⁶ establecida en 1919 por el Consejo Nacional de Investigación, la asociación científica internacional sin fines de lucro – con 60,000 miembros en 137 países– crea la *Declaración de la AGU*.

Esta declaración convierte a los datos en un patrimonio mundial, que deben ser conservados, pues ayudarán a los científicos a comprender los sistemas terrestre, planetario y heliofísico. Es necesario que sean conservados a largo plazo para

⁴⁶ Declaración adoptada por la *American Geophysical Union* el 29 de mayo de 1997; y fue refirmado en mayo de 2001, mayo de 2005, mayo de 2006 y finalmente revisado y refirmado en mayo de 2009, febrero de 2012 y septiembre de 2015.

futuro uso. Deben ser puestos a disposición de la comunidad científica y del público tan pronto como sea posible. Además, deben ser accesibles, en formatos utilizables, con suficiente documentación legible que permita su reutilización.

Los procesos de recopilación de información y análisis de los datos en las Ciencias de la Tierra resultan esenciales para la comprensión del entorno natural, pues cambian a cada momento (tiempo). A diferencia de otros campos científicos que están muy alejados de la vida de la mayoría de las personas, en esta área se tratan los procesos que influyen todos los días en el mundo –como las tasas de lluvia, tendencias en las pesquerías marinas, probabilidades de terremotos, potenciales de erupción volcánica, entre tantas otras–.

Es por eso que la adhesión a esta política fomenta los avances científicos y produce beneficios económicos que repercuten en la toma de decisiones, a fin de mejorar la seguridad pública y el bienestar social, mismos que contribuyen a la seguridad nacional y mundial al mantener a la sociedad más informada.

Asimismo, busca ser lo más abierta posible para maximizar el rendimiento científico, económico y social de la obtención de los datos, ya sea con fondos públicos y/o privados. Solo existirán situaciones limitadas de acceso por cuestiones éticas, pero no por propiedad. Estas deben ser justificadas a un nivel de protección de seguridad, la confidencialidad para así poder evitar daños a entidades protegidas. La importancia es documentar las tendencias y la producción de los datos a largo plazo para comprender muchos fenómenos naturales, debido a que el estado de los sistemas naturales nunca se repite, las pérdidas de datos o las oportunidades de recopilación no pueden ser corregidas.

La AGU es clara tanto en su postura como en su declaración, considerando el valor de los datos que crece con el tiempo, llevando a cabo ventajas en la curación de datos. Por lo tanto, los datos y la documentación asociada deben depositarse en un repositorio de la especialidad, siendo parte integral de los programas operativos y de la investigación, por lo que la comunidad científica tiene un amplio conjunto de datos para la realización de sus investigaciones en beneficio de la sociedad y de la ciencia.

Los principios rectores de FAIR para la gestión de datos científicos y la administración

La necesidad urgente de mejorar la infraestructura que respalda a la reutilización de datos académicos y el grupo de personas interesadas –que representan a la academia, la industria, las agencias de financiamiento y los editores académicos–, se unieron para diseñar y respaldar los Principios que referimos, con la intención de utilizarlos como una guía para aquellos interesados que deseen mejorar la reutilización de sus datos. A diferencia de las iniciativas de pares –que se centran en especialistas–, estos principios ponen especial énfasis en mejorar la capacidad de las máquinas para encontrar y utilizar los datos automáticamente, además de respaldar su reutilización por parte de los individuos.

Este proyecto desarrolla estándares que conectan a investigadores, editores y repositorios de datos para habilitar los datos (localizables, accesibles, interoperables y reutilizables) a gran escala. Apunta también a acelerar el descubrimiento científico y mejorar la integridad, transparencia y reproducción de los mismos.

Dentro de los principales, están los habilitadores como la AGU, re3data y DataCite, quienes decidieron desarrollar una nueva herramienta que apoye a los investigadores a encontrar un repositorio adecuado para sus datos, como el *Repository Finder*.⁴⁷ Este buscador de repositorios es un proyecto liderado por la Unión Geofísica Americana (AGU) en asociación con DataCite y la comunidad de Ciencias de la Tierra, el Espacio y el Medio Ambiente.

Los datos son accesibles, interoperables y reutilizables no solo para las personas, sino también para las máquinas con identificadores y metadatos únicos y persistentes, apropiados para ayudar al descubrimiento y ser citados de manera equivalente a otros resultados académicos, mediante un protocolo estándar basado en la web, proporcionando la información de procedencia y el licenciamiento de uso y sean interoperables técnica y semánticamente, así como reutilizables. Los datos también deben estar bien conservados, ser accesibles y estar vinculados de forma segura a las publicaciones asociadas, recursos o facilitar el reuso.

Han surgido estándares para el uso y tratamiento de los datos. A continuación se describe mediante el esquema que plantea Tim Berners-Lee,⁴⁸ el cual desarrolla la asignación de cinco estrellas para datos abiertos a través de cinco escalones o niveles:

1. *Una estrella*: publica tus datos en la Web (con cualquier formato) y bajo una licencia abierta. Significa que el público que encuentre los datos tendrá el beneficio de acceder libremente a la información bajo una licencia de uso.
2. *Dos estrellas*: los datos son publicados bajo una estructura con licencia abierta (*ei*: hoja de cálculo en vez de una imagen de una tabla escaneada), es decir, los datos deberán estar estructurados en un formato legible por máquina. Por lo que el usuario podrá reutilizar los datos publicados y hacer su propio análisis.
3. *Tres estrellas*: se deben usar formatos no propietarios (*ei*: CSV en vez de Excel), es decir, datos abiertos, estructurados, y sin propietarios. En este nivel los datos abiertos empiezan a consolidar su potencial para que su reutilización sea fácil, productiva, además de que genera impacto.
4. *Cuatro estrellas*: deben emplear URIs. Los datos deben integrarse con otras fuentes de datos y aplicaciones. Se pasa de datos abiertos a datos vinculados. Por lo que se facilita su identificación y localización mediante

⁴⁷ Datacite. *Repository Finder*. Disponible en <https://repositoryfinder.datacite.org/>. (Consultado: 25-01-19).

⁴⁸ Hausenblas, M. *Datos Abiertos*. Disponible en <https://5stardata.info/es/>. (Consultado: 25-01-19).

“identificador uniforme de recursos” o *URI*, reconociendo, por ejemplo, la “dirección” de un sitio en la red.

5. *Cinco estrellas*: enlaza tus datos a otros datos para proveer contexto. Por ello, los datos deben estar vinculados con otros datos. Al integrar los datos con otra información relevante, la fusión de estas dos o más fuentes de información pone todo este conocimiento en un nuevo contexto, creando valor adicional.

1.5.1. Antecedentes del Repositorio de datos

Desde el contexto de la comunidad científica, en específico en el área de las Ciencias de la Tierra, crear una herramienta como lo es un repositorio de datos plantea retos técnicos y sociales en donde existen iniciativas, directrices y políticas de agencias financiadoras, organismos gubernamentales y privados que son considerados para incentivar esta práctica.

La ciencia genera datos y a estos se les puede considerar la base fundamental de esta. Actualmente, se están creando entornos mucho más abiertos, buscando tener acceso a aquellos, mismos que se generan de manera constante. Esto cambia los hábitos de los investigadores, pues se generan nuevos perfiles de profesionales y nos adentra a un nuevo campo en diversas disciplinas.

Bajo este contexto se habla de la *e-ciencia*, que considera la práctica de compartir los datos de las investigaciones publicadas en Acceso Abierto mediante iniciativas que apoyan el impulso y la difusión de esta nueva postura de la comunidad científica.

Este movimiento busca hacer que la investigación científica y los datos sean accesibles para todos. Mediante diversas prácticas como la publicación, campañas para el Acceso Abierto, con la finalidad de crear nuevo y más conocimiento científico a través de canales de mayor transparencia y accesibles durante el proceso de investigación, incluyendo la ciencia de cuaderno abierto, la ciencia ciudadana y aspectos del software de código abierto y proyectos de investigación financiados.

Además, utiliza la prevalencia de Internet y de herramientas digitales para permitir una mayor colaboración de investigación local y global. Numerosos documentos, organizaciones y movimientos sociales abogan por una adopción más amplia de datos de ciencia abierta, fomentando el desarrollo y la implementación de estrategias de comunicación para la investigación científica, con el objetivo de propiciar la colaboración científica y el descubrimiento en nuevos campos.

Por ejemplo, en las Ciencias de la Tierra, los datos se muestran en una variedad de usuarios, tales como puntos de datos individuales, transmitidos desde una ubicación y/o franjas o cuadrículas de datos que cubren áreas geográficas. Ejemplos de ello los encontramos al tratar de satisfacer necesidades de datos homogéneos –como las diferencias de medición– y formato a gran escala –como

los datos satelitales administrados por las instalaciones de datos de la Administración Nacional de Aeronáutica y del Espacio (NASA)–.

En 1971, con la creación del *Protein Data Bank* (PDB), inicia un proyecto que buscaba compartir y distribuir datos bajo los auspicios de la American Crystallographic Association (Crawford, Hurd, Weller, 1996). Un banco de datos con crecimiento de 500 estructuras para sus investigaciones a finales de los ochenta y para 2010 ya contaba con más de 7,500,000 estructuras.

Bajo ese modelo surgieron otros repositorios similares en diversas disciplinas – como el *Neurodatabase*, que radicaba en el compromiso de acelerar el proceso científico de la *Data Observation Network for Earth* (DataONE), que satisfacía las necesidades de la Ciencia de la Tierra con el respaldo de la Fundación Nacional de Ciencias de los Estados Unidos de América.

Cada vez, se requiere que los datos de investigación estén disponibles, por lo que es necesario conocer el mundo científico y el intercambio de los mismos con el fin de hacer que los datos utilizados para la investigación académica estén disponibles para otros investigadores, haciendo uso del *data sharing*,⁴⁹ es decir, consiste en obtener y reusar los datos finales de la investigación entre los científicos para maximizar esfuerzos y recursos.

De esta manera, se logra un rendimiento y aprovechamiento de los recursos económicos que se invierten en la ciencia, financiados con fondos públicos. La libre disposición de los datos permitiría realizar nuevos estudios y generación de nuevo conocimiento, sobre todo facilitaría una mayor transparencia por parte de los investigadores.

La política de la *National Science Foundation*⁵⁰ (NSF) espera que los investigadores compartan con otros investigadores, sin costo adicional, los datos, muestras, colecciones físicas y otros materiales de apoyo, creados o recopilados, en el transcurso del trabajo. Y alienta a compartir software e invenciones que apoyen para llevar a cabo innovaciones útiles y utilizables.

Borgman⁵¹ dice: “[...] los datos son un concepto difícil de definir, ya que los datos adoptan muchas formas, tanto físicas como digitales”. Por otro lado, Pampel *et al.* definen que los datos de investigación “son una parte [descriptiva] del resultado de un proceso de investigación y abarca todas las etapas de la investigación. Y por lo tanto se producen en diferentes tipos de datos, niveles y formatos, de diferentes disciplinas métodos”.⁵²

⁴⁹ Vickers, J. (2006) Whose data set is it anyway? Sharing raw data from randomized trials. Disponible en <https://trialsjournal.biomedcentral.com/articles/10.1186/> (Consultado: 25-01-19).

⁵⁰ NSF. National Science Foundation Grant General Conditions (GC-1). Disponible en <https://www.nsf.gov/pubs/2001/gc101/gc101rev1.pdf>. (Consultado: 22-05-19).

⁵¹ Borgman C. L. “Who’s Got the Data? Interdependencies in Science and Technology Collaborations”. En *Springer Nature Switzerland AG*. Disponible en <https://link.springer.com/article/10.1007%2Fs10606-012-9169-z>. (Consultado: 25-05-19).

⁵² Pampel, H. , Vierkant, P., Scholze F. (2013) Making Research Data Repositories Visible. En PlosOne:

En Europa, el Grupo de trabajo del Proyecto RECOLECTA en conjunto con la Universidad Australiana de Melbourne dicen: “Los datos de la investigación son hechos, observaciones o experiencias que se basan en argumento, teoría o prueba y pueden ser numéricos, descriptivos o visuales, pueden encontrarse en estado bruto o analizado”.⁵³

Por otro lado, Dallmeier-Tiessen⁵⁴ dice que los datos de investigación pueden ser presentados a través de tres estrategias de publicación:

1. Como un objeto de información independiente, a través de un Repositorio.
2. Con documentación textual, llamado documento de datos.
3. Como enriquecimiento de una publicación de texto interpretativo.

Estas estrategias de publicación requieren una infraestructura de información, que garantiza el almacenamiento y la accesibilidad de los datos con un máximo de persistencia y confiabilidad. Dichas infraestructuras pueden incluir los llamados *archivos de datos, centros de datos, bibliotecas digitales, colecciones digitales y similares*.

1.5.2. Definiciones, objetivos y características del Repositorio de Datos

El Repositorio de Datos se enfoca en sus propios objetivos particulares, su diseño y desarrollo, audiencia, marco de tiempo, infraestructura, flujo de trabajo y productos, ya sea con datos homogéneos o heterogéneos. La comparación de los esfuerzos del Repositorio proporciona información sobre la gestión de datos y el diseño del sistema.

Los objetivos del Repositorio de Datos son:

- Recopilar, manejar y utilizar los datos e interactuar entre ellos.
- Preservar los datos de cualquier área del conocimiento.
- Contener metadatos que impliquen la integridad descriptiva y la disponibilidad de técnicas de validación.
- Ofrecer servicios de gestión de datos definidos por el almacenamiento, archivo, conservación y curación.

Desde el un punto de vista de la investigación, en donde se preparan los datos para un análisis, esta tiene dos fines: el conocimiento y las cuestiones sociales. Por otro lado, y basándonos en diversas definiciones, fuentes de investigación y en las estructuras que se trabajan en el área de las Ciencias de la Tierra, hacemos

<https://doi.org/10.1371/journal.pone.0078080> (Consultado: 25-05-19).

⁵³ FECYT. Grupo de trabajo de “depósito y gestión de datos en Acceso Abierto” del proyecto RECOLECTA, 2012.

⁵⁴ Dallmeier-Tiessen S.; Darby, R.; Gitmans, K.; Herterich, P.; Lambert, S. *et al.* (2012). *Resumen de los estudios, publicaciones teme datos en acceso abiert*. Disponible en <http://www.alliancepermanentaccess.org/wp-content/plugins/download-monitor/download.php?id=Summaryofthestudies%2Cthematicpublicationsandrecommendations>. (Consultado 21-04-2019).

la diferencia en los datos de investigación para los tipos de Repositorios, siendo estos:

1. **Repositorio de Datos de Investigación (RDR, Research Data Repositories).** Son parte esencial de la infraestructura de investigación de diferentes instalaciones, recursos y servicios relacionados, utilizados por la comunidad científica para realizar investigaciones de alto nivel en sus respectivos campos.

Es importante comentar que al compartir datos de investigación y otros materiales, terceros pueden evaluar el conocimiento académico basado en ellos y en los objetos relacionados. Esta disponibilidad de datos para la investigación y materiales de investigación fomenta la transparencia y la confiabilidad en los procesos, pues contribuyen en gran medida a compartir y reutilizar los datos de investigación mediante recomendaciones para una gestión adecuada de los mismos, además de que siempre deben estar disponibles para cumplir con los principios de integridad de la investigación.

Estas son algunas de sus características:

- Se realizan a través de la integración de datos e información recopilada durante proyectos, salidas de campo, muestreo, etcétera.
- Organizados bajo las normatividades e indicaciones de las directrices.
- Deben incluir una breve declaración que describa cómo y en qué términos se puede acceder a cualquier información de investigación.
- Cada organización de investigación tendrá políticas específicas y procesos asociados para mantener los datos de investigación financiados con fondos públicos y de las solicitudes de terceros para acceder a dichos datos.
- Por lo regular, son financiados con fondos públicos que no se generan en formato digital y se almacenarán, de manera que se compartan en caso de que se solicite un acceso válido a los datos que se reciben (esta expectativa podría cumplirse implementando una política para convertir y almacenar dichos datos en formato digital de manera oportuna).
- Cuando el acceso a los datos está restringido, los metadatos publicados también deben indicar el motivo y resumir las condiciones que deben cumplirse para que se otorgue el acceso.
- Las organizaciones de investigación se asegurarán de que se proporcione una curación de datos efectiva a lo largo de todo el ciclo de vida de los datos, con la “curación de datos” y el “ciclo de vida de los datos”, según lo definido por el Centro de Curación Digital

Por lo regular, este tipo de datos siempre se respaldan en una publicación o proyecto, como resultado de ello y también para llevar a cabo su conservación. Deben estar incluidos metadatos determinados por los especialistas de la información, así como de la comunidad de investigadores o del proyecto.

2. Repositorio de Datos para la Investigación (DRR, Data Repositories for Research). Tienen las condiciones y el formato para ser reutilizados con diferentes fines y propósitos científicos. Están programados para generar nuevo conocimiento, descubrir otras líneas de investigación, trabajar bajo una dinámica diferente y relacionar a grupos multidisciplinarios de especialistas de un área científica determinada.

Cambian la dinámica de trabajo y de investigación al procesar y ocuparse con este tipo de datos y dentro de las Ciencias de la Tierra, el Mar y la Atmósfera. Son datos que representan grabaciones de eventos o análisis y observación del estado de la Tierra, del Sistema Solar en tiempo real que nunca se pueden repetir, siendo estos datos nuevos y únicos. Cada vez más, estos datos, modelos y software suministran beneficios sociales de gran impacto, ya sea económico, de protección civil y de investigación. Dadas estas conexiones, reconocemos que es esencial garantizar la calidad, el valor y la integridad de los datos y otros recursos relacionados con las publicaciones académicas.

La publicación de artículos académicos en la comunidad científica de Ciencias de la Tierra está supeditada a la disponibilidad simultánea de los datos, por lo que deben ser compartidos, abiertos y almacenados en repositorios alineados con la comunidad.

¿Qué son los datos para la investigación? Se pensaría que son los generados por un proceso de investigación, realizado por una persona o un grupo de académicos que hacen experimentos o generan información con base en un análisis de contenidos y de resultados. No obstante, en una propuesta de este tipo el concepto cambia y se definen como:

La obtención de un conjunto de datos especializados sobre una línea de estudio, por medio de procesos y detección de eventos a través de tecnología de punta, que bajo parámetros establecidos y sistemas automatizados ofrecen datos en tiempo real que bajo una interpretación estructurada y metodológica mediante un análisis de contenido se convierten en información útil para la investigación.⁵⁵

Bajo estos conceptos, los *datos para la investigación* son de gran utilidad para la labor científica y académica de los investigadores, cualquiera que sea su línea, debido a que de los datos obtenidos pueden ofrecer diversas interpretaciones, generar otras líneas en diferentes disciplinas de la fuente creadora, realizar estudios o procesos de investigación que ayuden a la generación de nuevos conocimientos escritos, impresos, electrónicos o inductivos en beneficio de la ciencia y de la sociedad, estableciendo parámetros de estudio de acuerdo con los intereses y propuestas de desarrollo que se deseen cubrir.

⁵⁵ Definición estructurada por el personal de la Biblioteca Conjunta de Ciencias de la Tierra de la UNAM, con base en su experiencia en el manejo de datos de investigación y el desarrollo de Repositorios Institucionales, utilizando DSpace, Dublin Core y Open Aire.

Características

Estos datos cuentan con características que los hacen ser valorados como fuente primaria de información:

- **Únicos** por el hecho de que son generados por un sistema en tiempo real, de acuerdo con las necesidades de contenido y temática manejadas, bajo una estructura de información a obtener. No son repetibles o iguales al ser obtenidos de fenómenos con características distintas.
- **Constante** en su proceso de generación y no se interrumpen por períodos de tiempo, detectando o analizando los fenómenos que se presenten de manera regular sin descuidar la obtención de datos de cada uno de ellos para que se tenga un control de calidad y una información real de lo que se estudia.
- **Infinitos**. Lo que obliga a establecer un proyecto que genere datos de eventos o fenómenos que afecten de manera permanente a la sociedad y al mismo tiempo sean distintos en cada caso, de acuerdo con el tiempo de estudio y/o análisis.
- **Permiten la reutilización de datos** para investigaciones no previstas por los investigadores iniciales. Evita la duplicación innecesaria de la recopilación de datos.
- Se lleva a cabo la **creación de análisis de datos** mediante la combinación de datos de múltiples fuentes.

1.5.3. Tipos y alcances

El Repositorio de Datos se valora por la administración, el acceso a los datos, la mejora de la revisión por pares y la calidad de los productos digitales. Están respaldados y vinculados para garantizar su acceso a otros datos, software, servicios y otros productos de investigación digital relacionados. Asimismo, son tomados muy en cuenta por la comunidad científica al facilitar el acceso a los hallazgos de la investigación y ofrecer a su vez una mayor capacidad de información para descubrir e integrar diversos conjuntos de datos rápidamente para planificar la sostenibilidad o responder en tiempo real a desastres, entre muchas otras actividades más. En ese sentido, podemos dividir los Repositorios de Datos como se muestra a continuación.

Tipos de Repositorios de Datos en la investigación:

- *De Investigación Institucional*. Están a cargo de una institución –como una universidad o un centro de investigación–. A nivel universitario el ámbito es multidisciplinar, pero, sobre todo, permite el ingreso de múltiples datos, dependiendo las variables y los intereses de los académicos.
- *De Investigación Disciplinaria y Multidisciplinaria*. Se enfocan en un área temática específica operando por medio de infraestructura, según la naturaleza del dato. Estos Repositorios se dedican a almacenar, publicar y distribuir en torno a una investigación, permitiendo a los investigadores

publicar todos sus datos de manera citable, de búsqueda y de forma compatible.

- *De Investigación Específica del Proyecto*. Son de enfoque concreto, siendo el resultado de proyectos de investigaciones particulares.

En este ámbito científico, los investigadores están de acuerdo con el beneficio potencial del intercambio de datos para el avance en las investigaciones, pero aún no se libera en su totalidad. Hoy en día, en la mayoría de las disciplinas académicas es difícil obtener una visión general comparativa del RDR existente. Los sitios como el Directorio de Repositorios de Acceso Abierto (OpenDOAR) y Registro de Repositorios de Acceso Abierto ROAR solo contienen una pequeña parte que se refiere a la Investigación en Repositorios de Datos (<5%), ya que ambos registros se centran en Repositorios para publicaciones académicas o – como se conocen– de literatura académica. En los últimos años, sitios web como Open Access Directory (OAD) o DataCite comienzan a listar los Repositorios de Datos, pero con muy poca información sobre su descripción y servicios.

Repositorios e infraestructura de ecosistemas de datos

Dentro de los Repositorios de Datos se maneja el concepto de *ecosistema de datos*, que es clave, pues fomenta el pensamiento sobre la interrelación de una multiplicidad de Repositorios, así como las actividades asociadas con los datos en los ámbitos de investigación y Repositorio. Por lo que en este caso lo definimos como *una recopilación, manejo y utilización de los datos y la interacción entre ellos*.

Históricamente, gran parte de la planificación de los datos y los Repositorios de Datos se realizó de forma independiente, como en laboratorios, disciplinarias y empresas. Hoy en día, cada vez existen más solicitudes bajo el modelo de datos abiertos, con arquitectura de sistemas diseñados con base en las líneas de investigación que se estén desarrollando, por lo que debe considerarse el contexto cuando se reúnen los datos. El concepto de *ecosistema de datos* captura la dinámica y los comentarios asociados con los datos y los Repositorios de Datos, por lo que a continuación se presentan algunos ejemplos:

- *National Computational Infrastructure* (NCI). Diseñado para la recopilación y publicación de metadatos de Geonetwork, en distintos proyectos. Actualmente, almacena y administra más de 10PB de conjuntos de datos de importancia nacional e internacional. La recopilación de datos de cada colección tiene un catálogo de información de resumen y cuenta con un plan de investigación y de recolección de datos para su gestión. Ofrece apoyo y asesoría en la identificación y en la creación de un *Digital Object Identifier* (DOI) y asignación de un *Uniform Resource Identifier* (URI), etcétera. Disponible en <https://datamgt.nci.org.au/>.
- *AuScope*. Esta herramienta cuenta con un conjunto de herramientas de datos y análisis a investigadores, comprendiendo datos geocientíficos tanto gubernamentales como de la industria a través de varios programas

integrados para la investigación en sistemas de la Tierra y su impacto en la calidad de vida y el medio ambiente de Australia. Es de acceso libre y se puede consultar en <https://www.auscope.org.au/>.

- *Australian Research*. Lleva a cabo el Proyecto ARROW, que respalda los Repositorios Académicos Independientes (aquellos que no están asociados a instituciones). Una actividad complementaria es el desarrollo y servicios nacionales desarrollados por la Biblioteca Nacional, utilizando los metadatos recopilados para proporcionar servicios a través de protocolos. Disponible en <https://www.arc.gov.au/>.
- Data Commons (ARDC). Este sistema trabaja los datos de investigación y proyectos asociados, investigadores y servicios de datos. Se puede acceder a la reutilización de los datos para la investigación de más de cien organizaciones de investigación australianas, agencias gubernamentales e instituciones culturales. Incluye resultados de encuestas; datos recopilados automáticamente de programas informáticos, sensores o instrumentos; imágenes, audio y video; cuerpos de texto; páginas web y blogs. Disponible en <https://ardc.edu.au/>.
- Center for Open Science (COS). Desarrolla una variedad de herramientas de software, flujos de trabajo y soluciones de almacenamiento de datos llamadas OSF para ayudar a los investigadores a gestionar, archivar, descubrir y compartir investigaciones de forma más abierta. También apoya en la identificación, conexión en las comunidades de investigadores, instituciones afiliadas, financiadores y editores de los resultados de la investigación. Disponible en <https://cos.io/our-services/research/>.

Licencias de datos para los datos

Para poder establecer una comunicación clara y que se rija bajo las condiciones de uso adecuadas de los conjuntos de datos, se deben utilizar los permisos y licenciamientos que muestren los usos a los que pueden someterse los datos en caso de reutilización y las condiciones que deberán tener en cuenta en materia de propiedad intelectual. Entre estos licenciamientos tenemos:

- a) **Carta de responsabilidad de derechos**. Documento por el cual se otorga la licencia no exclusiva y sin límite de temporalidad el uso, publicación y distribución de los datos la cual se hace extensiva en formatos impresos y en cualquier otro medio electrónico. Sin finalidad lucrativa, sino académica, con otorgamiento y autorización correspondiente para que la difusión sea transparente y pueda efectuarse a través de medios electrónicos, tanto en red local como por vía Internet.
- b) **Licencia Pública General (GNU) (General Public License)**. Es una licencia de derecho de autor ampliamente usada en el mundo del software libre y código abierto. Garantiza a los usuarios finales (personas, organizaciones, compañías) la libertad de usar, estudiar, compartir (copiar) y modificar el software. Su propósito es doble: declarar que el software cubierto por esta licencia es libre, y protegerlo (mediante una práctica conocida como

Copyleft) de intentos de apropiación que restrinjan esas libertades a nuevos usuarios cada vez que la obra es distribuida, modificada o ampliada.

- c) **CC (Creative Commons)**. Esta organización sin fines de lucro promueve el acceso y el intercambio de cultura y desarrolla instrumentos jurídicos de carácter gratuito que facilitan usar y compartir creatividad y conocimiento. Sus licencias cuentan con tres capas: código legal, legibilidad por humanos y maquinas. Y se componen por cuatro módulos:
1. *Attribution / Atribución (BY)*. Requiere la referencia al autor original.
 2. *Share Alike / Compartir Igual (SA)*. Permite obras (objetos digitales) derivadas bajo la misma licencia o similar (posterior u otra versión por estar en distinta jurisdicción).
 3. *Non-Commercial / No Comercial (NC)*. Obliga a que no sea utilizada con fines comerciales.
 4. *No Derivative Works / No Derivadas (ND)*. No permite modificación de ninguna manera.

Y se combinan para dar lugar a las seis licencias:

1. Attribution / Atribución (CC BY).
2. Attribution Share Alike / Atribución Compartir Igual (CC BY-SA).
3. Attribution NoDerivatives / Atribución No Derivadas (CC BY-ND).
4. Attribution Non-Commercial / Atribución No Comercial (CC BY-NC).
5. Attribution Non-Commercial Share Alike / Atribución No Comercial-Compartir Igual (CC BY-NC-SA).
6. Attribution Non-Commercial No Derivatives / Atribución No Comercial-No Derivadas (CC BY-NC-ND).

De reciente creación CCSearch (disponible en <https://search.creativecommons.org/>), herramienta que permite que los trabajos de dominio público y con licencia abierta sean buscados y recuperados. Se ha utilizado más de 1,400 millones de veces para ayudar a los creadores a compartir conocimientos y creatividad en línea; busca en más de 300 millones de imágenes desde API abiertas.

- d) **Open Data Commons**. Prepara un conjunto de licencias adecuadas para los diferentes tipos de datos que se manejan, dependiendo los niveles de consulta y difusión.
- e) **Public Domain Dedication and License (PDDL)**. Es un documento destinado a permitirle al autor compartir, modificar y utilizar libremente este trabajo para cualquier propósito y sin restricción alguna. Esta licencia está diseñada para su uso en bases de datos o sus contenidos (datos), ya sea de forma conjunta o individual.
- f) **Attribution License (ODC-By)**. Pone el material en el dominio público para que los usuarios compartan, modifiquen y utilicen esta base de datos de forma gratuita.
- g) **Open Database License (ODC-ODbL)**. Es un acuerdo de licencia destinado a permitir que los usuarios compartan, modifiquen y utilicen

esta base de datos de forma gratuita y mantengan esta misma libertad para los demás.

- h) **OGL (Open Government License)**. Recomienda que use y reutilice la información de manera libre y flexible, con pocas condiciones. Otorga una licencia mundial, libre de regalías, perpetua, no exclusiva para utilizar la información sujeta a las condiciones que se detallan a continuación. Y no afecta su libertad bajo trato justo o uso justo o cualquier otra excepción y limitación de derechos de autor o de base de datos. Por lo que es libre de:
- Copiar, publicar, distribuir y transmitir la información.
 - Adaptar la información.
 - Explotar la información comercialmente y no comercialmente; por ejemplo, combinándola con otra información o incluyéndola en su propio producto o aplicación.
- i) **BSD (Distribución de Software de Berkeley)**. Es la licencia de software otorgada principalmente para los sistemas BSD (Berkeley Software Distribution). Tiene como requisitos la reciprocidad de compartir-igual y menos restricciones en comparación con otras como la GPL, estando muy cercana al dominio público. La licencia BSD, al contrario que la GPL, permite el uso del código fuente en software no libre. La versión original ya se ha revisado y sus variantes son denominadas *licencias BSD modificadas*.
- j) **Licencia MIT**. Se origina en el Instituto Tecnológico de Massachusetts. Es un software libre permisivo, lo que significa que impone muy pocas limitaciones en la reutilización; por tanto, posee una excelente compatibilidad de licencia. Permite reutilizar software dentro del software propietario. Por otro lado, la Licencia MIT es compatible con muchas licencias *copyleft*, como la GNU General Public License (software con licencia MIT puede integrarse en software con licencia GPL, pero no al contrario).
- k) **Licencia Apache 2.0**. Es de software libre, permisiva, creada por la *Apache Software Foundation* (ASF). Requiere la conservación del aviso de derecho de autor y el descargo de responsabilidad, pero no es una licencia *copyleft*, ya que no requiere la redistribución del código fuente cuando se distribuyen versiones modificadas.

Servicios

Se pretende sean los más adecuados y respalden la gestión, el almacenamiento, el acceso y el uso de los datos de investigación, como los Repositorios de Datos de Investigación, los portales y otros proveedores de servicios. Entre ellos se ofrece:

- Permitir el acceso universal y equitativo a datos científicos, servicios de datos, productos e información con garantía de calidad.
- Asegurar la administración de datos a largo plazo.
- Fomentar el cumplimiento de los estándares y convenciones de datos acordados.

- Proporcionar mecanismos para facilitar y mejorar el acceso a datos y productos de datos.

Además, los datos deben estar disponibles en diversas plataformas, como texto sin formato, imágenes, datos sin procesar, gráficos estructurados, texto estructurado, datos archivados, aplicaciones de software, datos audiovisuales, bases de datos, datos basados en la red, código fuente, entre muchos otros.

1.5.4. Repositorio de Datos en Ciencias de la Tierra

En 1997, la American Geophysical Union (AGU) inició con la administración de los datos generados por sus miembros y a desarrollar una de las primeras declaraciones sobre su posición de la iniciativa en torno a los datos abiertos. En ella buscaba se reconociera debidamente y documentara a los futuros científicos de datos (datos para la comprensión de los sistemas terrestre, planetario y heliofísico, entre otros) y garantizara su preservación. Por lo que Christine McEntee, Directora Ejecutiva (CEO) de AGU comentó:

Hoy, con el generoso apoyo de la Fundación Laura y John Arnold, nuestra comunidad está trabajando en conjunto para garantizar que las Ciencias de la Tierra y el Espacio, incluidas más de 50,000 publicaciones, sean el primer campo científico en tener datos abiertos y bien descritos.

También se encuentra trabajando dentro de este contexto la Coalición para la Publicación de Datos en las Ciencias de la Tierra y el Espacio (COPDESS), que reúne a las principales editoriales, instituciones u organismos que generan datos, así como consorcios en las Ciencias de la Tierra y el Espacio con el objetivo de apoyar en la implementación conjuntamente y promover políticas y procedimientos comunes para la publicación y las citas de datos a través de revistas de esta área.

En 2014, esta Coalición adoptó una *Declaración de Compromiso* con respecto al apoyo y el seguimiento de las mejores prácticas en torno a los datos abiertos que incluía objetivos para la publicación en revistas, mismas que requieren que los datos que respaldan las conclusiones publicadas estén disponibles y se depositen, siempre que sea posible, en los Repositorios de Dominio principales para que los repositorios puedan desarrollar prácticas que respalden a los editores, y exista la disponibilidad y calidad de los datos. Así, y para ambos, se apoya una infraestructura que mejoraría la curación de datos y la integridad general en publicaciones académicas.

En el 2017, la AGU reunió a representantes de la comunidad internacional de Ciencias de la Tierra, el Espacio y el Medio Ambiente para trabajar en el proyecto de habilitación de datos FAIR. En este se desarrollan estándares que conectarán investigadores, editores y repositorios de datos (localizables, accesibles, interoperables y reutilizables) a gran escala. Por lo que apunta hacia un

aceleramiento a generar nuevo conocimiento científico y mejorar la integridad, transparencia y reproducción de los datos.

Es por eso que los datos que se generan en el área de las Ciencias de la Tierra y el Espacio son patrimonio mundial y deben ser acreditados, preservados y abiertos bajo una responsabilidad integral de los científicos –administradores de datos– e instituciones patrocinadoras. Son patrimonio mundial debidamente documentado, acreditado y preservado, que ayuda a futuros científicos a comprender La Tierra, los sistemas planetarios y la Heliofísica, debiendo estar disponibles para la comunidad científica y el público tan pronto como sea posible, en formatos accesibles y utilizables con suficiente legibilidad por máquina.

A fin de fomentar una política que apoye los avances científicos y produzca beneficios económicos y mejore la toma de decisiones, la seguridad y el bienestar de las naciones y contribuya a la seguridad nacional y mundial, manteniendo una sociedad bien informada, los datos de este tipo deben ser protegidos y difundidos.

En cuanto a la publicación de artículos académicos en la comunidad científica de La Tierra, el Espacio y el Medio Ambiente, está supeditada a la disponibilidad simultánea de los datos que respaldan el hallazgo de la investigación pues estos deben estar disponibles en la mayor medida posible, con posibilidad de ser compartidos, abiertos y almacenados en Repositorios alineados con los principios FAIR aprobados por la comunidad.

Las comunidades de datos científicos en el área de Ciencias de la Tierra y el espacio

Este grupo de científicos se esfuerza por convertirse en una comunidad mundial de excelencia en la creación y manejo de datos científicos mediante directorios, catálogos y repositorios que aseguren la administración a largo plazo y el suministro de datos de calidad y servicios de datos a la comunidad científica internacional.

Entre las instituciones representativas de este grupo de científicos tenemos a:

- La *American Geophysical Union (AGU)*, dedicada a promover las Ciencias de la Tierra y el Espacio para el beneficio de la Humanidad a través de sus publicaciones académicas, conferencias y programas de divulgación. Esta es una organización científica profesional sin fines de lucro que representa a cerca de 60,000 miembros en 139 países.
- La *European Geosciences Union (EGU)*, sin fines de lucro en los campos de las Ciencias de la Tierra, planetarias y espaciales. Con sede en Múnich (Alemania). Publica revistas científicas de Acceso Abierto y su evento más destacado es la Asamblea General, que reúne a más de 14,000 científicos de todo el mundo en el área de la Vulcanología, la exploración planetaria, la estructura interna de La Tierra y la Atmósfera, el cambio climático y las energías renovables.

- La *Earth Science Information Partners (ESIP)* es la comunidad de profesionales de tecnología de datos e información que se unen para coordinar los esfuerzos de interoperabilidad de las Ciencias de la Tierra. Permite a los miembros mejorar sus capacidades de gestión de datos. Creada por la NASA en 1998, incluye más de 120 organismos miembros, así como centros de datos federales, laboratorios gubernamentales de investigación, entidades de investigación de universidades, proveedores de recursos educativos, desarrolladores de tecnología y empresas sin fines de lucro y comerciales.
- La *Research Data Alliance (RDA)*. Comunidad de investigación que inició en 2013 gracias a la Comisión Europea, la American National Science Foundation, el Instituto Nacional de Estándares y Tecnología y el Departamento Australiano de Innovación. Construye los canales sociales y técnicos para el intercambio abierto de datos y que los investigadores compartan libremente datos a través de tecnologías, disciplinas y países para enfrentar los grandes desafíos de la sociedad.
- El *EarthCube / Council for Data Facilities* es una comunidad de científicos en el área de Geociencias y Geoinformática y científicos de datos. Participa en el proyecto de la Dirección de Geociencias de la NSF y la División de Infraestructura Cibernética Avanzada para transformar la investigación y mejorar el acceso, intercambio y el despliegue de los datos, así el como análisis de todas las formas de datos y recursos relacionados.
- La *Coalition on Publishing Data in Earth and Space Sciences (COPDESS)*, (Coalición para la Publicación de Datos en las Ciencias de la Tierra y del Espacio). Conecta a los editores de Ciencias de la Tierra y el Espacio y las instalaciones de datos para ayudar a traducir las aspiraciones de datos abiertos, disponibles y útiles de la política a la práctica. Ha desarrollado una declaración de compromiso, firmada por la mayoría de los principales editores y repositorios, proporcionando un directorio con las mejores prácticas recomendadas sobre datos e identificadores.
- El *FORCE11*. Comunidad de académicos, bibliotecarios, archivistas, editores y financiadores de investigación que ha surgido orgánicamente para ayudar a facilitar la creación e intercambio del conocimiento por medio de las comunicaciones académicas modernas y fomentando la publicación digital mediante herramientas de software y la difusión de la investigación a través de los medios de comunicación.
- La *National Oceanic and Atmospheric Administration* (Administración Nacional Oceánica y Atmosférica) (NOAA) es una agencia científica del Departamento de Comercio de los Estados Unidos, cuyas actividades se centran en las condiciones de los océanos y la atmósfera. Emite advertencias sobre condiciones meteorológicas peligrosas, prepara cartas de mares y de cielos, guía sobre el uso y la protección de los recursos oceánicos y costeros y conduce estudios para mejorar el entendimiento y la administración del ambiente.
- El *Grupo de Investigación Datasea*, conformado por investigadores del Valencia International Campus of Excellence (UPV, UV y CSIC). Tiene

como objetivo fomentar el intercambio de los datos científicos entre los investigadores. Desarrolló la Web Datasea, la cual es un portal enfocado a profesionales, investigadores, editores y gestores de la investigación interesados en el Acceso Abierto a los datos, en su uso compartido y reutilización de los mismos. Como investigador, profesional, editor, se puede interactuar con las herramientas y recursos. Disponible en <http://www.datasea.es/es/>.

Como podemos ver, las asociaciones, uniones, federaciones, agencias y grupos de trabajo de las áreas de las Ciencias de La Tierra, el Mar y la Atmósfera son pioneros en el manejo de datos para la investigación, por ser información especializada de apoyo a casi todas las demás ciencias. Las estructuras que han formado permiten conocer ahora millones de datos de distintas líneas y sistemas automatizados que ofrecen la posibilidad de ser tratados para generar nuevo conocimiento y al mismo tiempo generar datos de investigación. Sus propuestas en el manejo libre pero también confidencial de los datos han abierto diferentes caminos para su manejo por medio de distintos sistemas automatizados, como los repositorios. Muestran una forma distinta de entender la disponibilidad de los datos que generan, los diferentes organismos –tanto públicos como privados–, siendo estos algunos de los pioneros que entraron al proyecto de ciencia abierta, pero con responsabilidad en su manejo y en su divulgación.

En ese sentido, la presente investigación se enfoca en los siguientes capítulos en examinar los Repositorios Institucionales de Datos para la investigación, desarrollados en la UNAM, y ofrecer la experiencia que se tiene sobre el tema gracias a la participación directa en su conformación.

CAPÍTULO DOS

Desarrollo de Repositorios de Datos Científicos

La disponibilidad de flujos de trabajo para la publicación de datos tiene un impacto muy sensible en los científicos, así como en sus prácticas de investigación y en las estrategias de financiación, debido a que muchas veces desconocen cómo dar a conocer lo que generan por medio de plataformas o consideran que los datos son de su propiedad. Actualmente, se requiere el acceso a los datos como resultado de una investigación o publicación de un artículo y también se necesitan poner a disponibilidad de los especialistas los datos duros y puros que sirvan para generar nuevo conocimiento u otras líneas de investigación, sin que ellos se encuentren disponibles u hospedados en los Repositorios correspondientes a su disciplina, fuente de financiación u otros dominios.

Podemos encontrar datos que se utilizan como fuentes primarias para soportar una investigación técnica o científica, de beca o una actividad artística que, por lo regular, se utilizan como evidencia de un proceso y son los datos más comúnmente aceptados. Pero podemos partir de otra premisa en donde los datos nacen a cada momento y no se interrumpen y aún no pasan por un análisis, uso y/o interpretación.

Es importante considerar que los contenidos digitales y no digitales tienen el potencial de convertirse en una investigación. Datos que pueden ser experimentales, de observación, operacionales, de terceros, sector público, seguimiento, procesados e incluso reutilizables, los cuales emergen de un proyecto de medición o de una investigación que genera una cantidad importante de datos para estar vigentes.

Para entender mejor esto, partimos de la definición *Open Data*, según *Open Knowledge Foundation* (2015), que dice: “El conocimiento está abierto si alguien puede acceder, utilizarlo, modificarlo y compartirlo libremente. Sujeto a medidas que preserven la procedencia y la apertura”.

Este concepto es muy específico y hace referencia a dos tipos de datos:

1. Los datos que están legalmente abiertos, lo que en la práctica significa que están publicados bajo una licencia abierta y que las condiciones de reutilización están limitadas a la obligación de citar la autoría de los mismos y bajo algunas condiciones de uso.
2. Los datos técnicamente abiertos, que significa que sus archivos pueden leerlos las computadoras y que están hechos en formatos no propietarios. Es decir, que están accesibles para todo el que lo solicite y que el formato de los archivos y sus contenidos no están restringidos a ningún software propietario en particular.

En los últimos años, varias organizaciones científicas nacionales e internacionales han emitido declaraciones y políticas que subrayan la necesidad de un inmediato almacenamiento, manejo, uso y reuso de los datos, permitiendo así su intercambio bajo la dinámica de utilizar, analizar y usarlos.

Para ampliar más este panorama se revisaron cuatro declaraciones, considerando que son las que más impacto y eco han tenido en el medio científico y, sobre todo, que están a favor del *open data*, además de resaltar las condiciones, políticas e indicaciones de uso que se le deben dar a los datos.

La primera de ellas denominada los *Principios de Bermudas (Bermuda Principles drafted for Human Genome Project free data Access)*, los cuales fueron redactados para el acceso gratuito a los datos del Proyecto Genoma Humano, en 1996, se acordó que toda la información de la secuencia genómica humana generada por los centros financiados debería estar disponible de forma gratuita y en dominio público con el fin de fomentar la investigación y el desarrollo, maximizando los beneficios para la sociedad, y, sobre todo, hacer que los datos experimentales estén disponibles solo después de su publicación. Estos principios reformaron las prácticas de una industria estableciendo la publicación rápida de datos que aún se encuentran en prepublicación.

La segunda declaración que se cita es la del *Acuerdo de Fort Lauderdale*, realizada en Florida. Los científicos y profesionales en biomedicina respaldan el uso libre y sin restricciones de los datos de secuenciación del genoma por parte de la comunidad científica antes de que esos datos se utilicen para su publicación. En este Acuerdo también participó el *Wellcome Trust*, a partir de 2003, y el principal resultado fue un sistema compartido de responsabilidad entre las agencias de financiamiento, los productores de recursos y los usuarios para mantener y ampliar un recurso comunitario de datos genómicos. Más tarde, los principios del Acuerdo fueron adoptados por las agencias de financiamiento pertinentes, incluido el Instituto Nacional de Salud.

Ambas declaraciones influyeron en gran medida en los acuerdos posteriores para el intercambio de datos y los modelos de Acceso Abierto que se manejarían después.

Para 2008, las políticas que apoyan el intercambio rápido y abierto de datos genómicos impulsan directamente el ritmo acelerado de las investigaciones a gran escala. Esta comunidad comienza a implementar políticas e infraestructura análogas para hacer que los datos de Proteómica a gran escala estén también ampliamente disponibles sobre una base precompetitiva.

En ese mismo año, el *Instituto Nacional del Cáncer (NCI)*, por sus siglas en inglés) convocó a la *Cumbre Internacional sobre la Política de Divulgación e Intercambio de Datos de Proteómica*, en Ámsterdam, Países Bajos, para identificar y abordar posibles obstáculos para un acceso rápido y abierto a los datos. Los seis principios acordados por las partes interesadas en la cumbre fueron: 1) Tiempo, 2) Amplitud, 3) Formato, 4) Disposición en Repositorios, 5) Medidas de calidad y 6) Responsabilidad de los Datos de Proteómica.

En un momento en que los avances de las tecnologías de secuenciación de ADN aumentaban y significan que muchos más laboratorios podían producir conjuntos

de datos masivos, y cada vez era mayor el número de campos (más allá de la secuenciación del genoma) y se enfrentan a sus propias políticas de intercambio de datos. En el 2009, en Toronto, se lleva a cabo un Taller de Lanzamiento de Datos, por Genome Canadá, y otras agencias de financiamiento. En donde se reúnen científicos, especialistas en ética, abogados, editores de revistas y representantes de fondos para alentar la publicación rápida de datos de pre-publicación para proyectos con los siguientes atributos: gran escala (que requiere recursos significativos), amplia utilidad, creación de conjuntos de datos de referencia y asociado con la compra comunitaria y trabajando en cuatro ejes:

1. Las agencias de financiamiento, las cuales deben facilitar la especificación de políticas de liberación de datos para proyectos.
2. Los productores de datos, que deben declarar sus intenciones y permitir el análisis de sus datos.
3. Los analistas/usuarios deben analizar libremente los datos publicados antes de la publicación y actuar con responsabilidad en la publicación de los análisis.
4. Los editores de revistas científicas, quienes deben involucrar a la comunidad de investigación sobre temas relacionados con la publicación de datos y brindar orientación a los autores y revisores sobre el uso de terceros de los datos de prepublicación en los manuscritos.

Para 2015, con la *Declaración de La Haya*, sobre el descubrimiento del conocimiento en la era digital –que pretende mejorar el acceso a la información, los datos y las ideas para el descubrimiento del conocimiento en la era digital mediante la eliminación de barreras para acceder a la riqueza de los datos producidos por la sociedad, permitiendo encontrar respuestas a los grandes retos de cualquier tipo, que se plantean en el mundo globalizado. Basada en cinco principios potenciales, tanto económicos como sociales, siendo estos:

1. La propiedad intelectual no se diseñó para regular el flujo libre de hechos, datos e ideas, pero tiene un objetivo clave: la promoción de la actividad de investigación.
2. Las personas deben tener la libertad para analizar y seguir la curiosidad intelectual sin temor de persecución o repercusiones.
3. Las licencias y los términos del contrato no deben restringir a los individuos de la utilización de hechos, datos e ideas.
4. Ética alrededor del uso de contenidos, las técnicas mineras deben continuar para evolucionar en respuesta al cambio de tecnología.
5. La innovación y la investigación comercial basadas en el uso de hechos, datos e ideas no deben ser restringidas por la Ley de Propiedad Intelectual.

El documento político *Carta de datos abiertos y anexo técnico del G8*, publicado el 18 de junio de 2013, considera este movimiento global de los datos abiertos en un lugar crucial que permite a las personas y organizaciones desarrollar nuevos conocimientos e innovaciones que pueden mejorar la vida de los demás y ayudar en el flujo de información dentro y fuera de los países. Si bien los gobiernos y las

empresas recopilan una amplia gama de datos, no siempre se comparten de manera que el público pueda descubrirlos, utilizarlos o comprenderlos fácilmente.

Los datos abiertos pueden aumentar la transparencia sobre lo que están haciendo el gobierno y las empresas y aumentan la conciencia acerca de cómo se utilizan los recursos naturales de los países, cómo se gastan los ingresos de las actividades extractivas y cómo se tramitan y gestionan las tierras.

Por otra parte, están los datos gubernamentales de libre disponibilidad que se pueden utilizar de maneras innovadoras para crear herramientas y productos útiles que ayuden a las personas a navegar en la vida moderna más fácilmente.

El G8 está de acuerdo en que los datos abiertos son un recurso sin explotar con un gran potencial para alentar la construcción de sociedades más fuertes e interconectadas que satisfagan mejor las necesidades de los usuarios y permitan que la innovación y la prosperidad florezcan. Por lo tanto, acordaron en seguir un conjunto de principios que mandatan que los datos sentaran la base para el acceso, la divulgación y su reutilización, siendo estos:

1. Los datos deben ser abiertos por defecto.
2. Oportunos y exhaustivos, además de contar con la calidad y cantidad necesaria.
3. Deben ser utilizables por todos y hacerlos accesibles.
4. Comparables e interoperables y liberar datos para mejorar la gobernabilidad.
5. Para el bien de la gobernanza y participación ciudadana.
6. Y liberar los datos para el desarrollo incluyente y de innovación.

Dicho documento se centra en los datos abiertos de los gobiernos, otras organizaciones de la sociedad civil y del sector privado, que son invitados a adoptar estos principios.

Es importante mencionar que dentro del desarrollo de un Repositorio de datos, se debe considerar un elemento esencial de apoyo a las publicaciones académicas debido a que es una fuente utilizada por la comunidad científica para dar a conocer los avances de sus investigaciones, siendo por lo regular, que los datos se complementan, respaldan y/o vinculan con la publicación (artículo, informe, libro, capítulo, tesis, etc.) o un proyecto ofreciendo como resultado su presentación.

Con el fin de llevar a cabo el repositorio para manejar y hospedar datos, se pueden considerar una serie de buenas prácticas en su desarrollo y en el uso de los datos, dentro de los cuales podemos destacar:

- a) *Citación y procedencia en publicaciones académicas* –cuando se considere necesario–. Cuando los investigadores utilizan datos creados por otros,

estos 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* 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 set 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; información geográfica de los datos; metodología y enlace a publicaciones y otra documentación; unidades de medida, protocolos, abreviaciones, códigos, símbolos asociados a los datos; licencia de uso; citación recomendada.
- e) Capacidad de vinculación de los datos abiertos, siempre que sea posible estar vinculados con otros datos basados en su contenido y contexto, con el fin de maximizar el valor semántico.

Dentro de los elementos primordiales que deben cubrirse en el Repositorio que albergará datos para su eficiente funcionamiento, presentamos algunos de ellos, según la revisión de una serie de documentos y con base en la experiencia que se ha tenido en este desarrollo:

- Bajo su estructura deben ser un medio eficaz para publicar datos de investigación.
- Deben asegurar la preservación de datos a largo plazo, contando de esta manera con una preservación de la información.
- Asignar DOI (identificadores) a los dataset bajo la política en Acceso Abierto, con miras a contar con un registro real de cada elemento incluido.

- La plataforma debe cumplir con requerimientos de agencias financiadoras y/o editores.
- Incrementa la visibilidad de los datos de investigación y, como consecuencia, de los proyectos en que encuadran en su contenido.
- No es una plataforma para albergar *big data*, sino que se centra en la llamada *long tail of science*⁵⁶.

A continuación se adjunta ejemplo detallado de una estructura de un archivo *DATASETNAMEreadme.txt*, que puede ser consultado en el **Anexo 2** de este trabajo.

Con la finalidad de reforzar esta información y hacer más visible los elementos señalados, se muestran algunos ejemplos de los datos que ya se encuentran hospedados en los Repositorios:

- ZENODO. Disponible en <https://zenodo.org/record/155809#.V-5C5PQsC3t>
- 4TU Centre for Research Data. Disponible en <http://data.4tu.nl/repository/uuid:54ed4c8f-e7b6-4139-bc20-dc8168c2f890>
- Boston Area Research Initiative Dataverse. Disponible en <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/1J0IBN>
- Figshare. Disponible en https://figshare.com/articles/TCGA_Pan-Cancer_expression_and_mutation_data_for_Project_Cognoma/3487685
- Data.gouv.fr. Disponible en <https://www.data.gouv.fr/fr/datasets/cancer-de-la-prostate-30-dinnovations-et-apres/>

Podemos decir que es necesario administrar y añadir valor a los datos producidos por la ciencia para facilitar su integración en proyectos o colecciones en los cuales puedan usarse. En ese sentido se manejan dos líneas en este análisis del uso de los datos: el primero, el enfoque de literatura e-Ciencia, que se dirige a la ingesta, archivo y entrega; por otro lado, la planificación de datos que va encaminada a la toma de muestras, a partir de datos de archivos para emplear y reutilizar.

En cuanto a los datos de Ciencias de La Tierra, que es la especialización que hemos tenido, a menudo son utilizados por una gran cantidad de investigadores académicos y gestores que se basan en los datos recogidos de diversas maneras sobre una amplia gama de temas y pueden beneficiarse enormemente de una gestión más prudente de los datos.

Nos atrevemos a decir que en los últimos años la investigación científica ha evolucionado a pasos agigantados gracias al movimiento *open access* y ahora la e-Ciencia está modificando en gran medida las prácticas de los investigadores y favoreciendo los flujos e intercambio de información, proceso ya casi habitual

⁵⁶ En este caso nos referimos a la cantidad inmensa de productos científicos generados, los cuales deben incluirse en este tipo de proyectos; por lo tanto, se tiene que hacer una análisis preciso, una correcta administración y gestión estructurada que nos permita incorporar la ciencia conforme a los intereses y necesidades de la comunidad científica.

entre los investigadores, pero sí es un hecho que los servicios de alojamiento de información ya no se están haciendo de forma local sino se están hospedando en servicios de nube: Dropbox, OneDrive, Google Drive, redes sociales, Academia.edu o Research Gate. Es por eso que actualmente nos encontramos en un momento en el que los investigadores muestran una actitud más abierta para difundir sus investigaciones, motivados en gran medida por las políticas *open access*, la creación de repositorios y las herramientas de difusión de la ciencia.

Pautas a las reglas de Acceso Abierto a publicaciones científicas y Acceso Abierto a datos de investigación en Horizonte 2020

El programa nace para apoyar la implementación de la estrategia *Europa 2020* y la iniciativa emblemática de *Unión por la Innovación*, contribuyendo directamente a abordar los principales retos de la sociedad, crear y mantener el liderazgo industrial en Europa, así como reforzar la excelencia de la base científica, esencial para la sostenibilidad, prosperidad y el bienestar de Europa a largo plazo.

La Comisión Europea anunció que en 2014 lanzaría su nuevo programa de financiación para investigación e innovación *Horizonte 2020*, con un presupuesto de 80,000 millones de euros, con el cual se presentarían las prioridades políticas de la estrategia Europa, que aborda las principales preocupaciones compartidas por sus ciudadanos y en otros lugares del mundo.

El programa se centra en tres Pilares:

- i. Ciencia excelente, reforzar la excelencia científica de la Unión a nivel mundial, principalmente en iniciativas de temática abierta y, en general, en proyectos individuales.
- ii. Liderazgo industrial, acelerar el desarrollo de las Tecnologías de la Información y la Comunicación (TIC), nanotecnología, materiales avanzados, biotecnología, fabricación y tecnología espacial para ayudar a las PyME innovadoras europeas a convertirse en empresas líderes en el mundo y facilitar la financiación de riesgo en actividades de investigación e innovación en su llegada al mercado.
- iii. Retos sociales, para aportar una respuesta directa a las prioridades políticas y los retos identificados en la estrategia *Europa 2020*, tales como la seguridad, la energía, el transporte, el cambio climático y el uso eficaz de los recursos, la salud y el envejecimiento, los métodos de producción respetuosos del medio ambiente y la gestión.

Su base es primordialmente el fortalecimiento de la investigación europea, abordando los desafíos sociales y cerrar las brechas entre la investigación, el mercado y la sociedad. El taller europeo de 2013 se centró en desafíos identificados por *Horizonte 2020*: la salud, el cambio demográfico y el bienestar. Estas tres áreas comprenden una gama muy amplia de temas sociales, que incluyen el diseño de sistemas de salud eficaces y rentables; políticas y estrategias para un envejecimiento feliz y saludable; y componentes más amplios

del bienestar, que se encuentran, por ejemplo, en las dimensiones ecológica, ambiental y educativa.

El nuevo desarrollo en *Horizonte 2020* es el *Programa Piloto de Datos de Investigación Abierta*, que busca mejorar y maximizar el acceso a los datos de investigación generados por proyectos financiados por la UE. Adoptando un enfoque activo para la gestión de datos de investigación, los que incluyen mayor velocidad y facilidad de acceso, eficiencia (se financia una vez que se han reutilizado los datos varias veces) y se mejora la calidad y la transparencia de la investigación. Tiene dos principales pilares: desarrollar un plan de gestión de datos (DMP)⁵⁷, y proporcionar el Acceso Abierto, sí es posible.

Las condiciones que se deben cumplir son las siguientes para el DMP:

- Desarrollar y mantener actualizado el Plan de Gestión de Datos. Este debe manejar los datos de investigación durante y después de un proyecto, así como identificar las acciones y estrategias que garanticen que los datos sean de calidad, seguros, sostenibles, accesibles y reutilizables, mediante el Acceso Abierto, en todo lo posible, cumpliendo con los requisitos académicos, legales y éticos.
- Depositar sus datos en un Repositorio de Datos de Investigación, mediante sus diferentes procesos.
- Asegurarse de que terceros puedan acceder, explotar, reproducir y difundir sus datos libremente.
- Proporcionar información relacionada que identifique o, en su defecto, facilite las herramientas necesarias para utilizar los datos sin procesar para validar su investigación.

Este piloto aplica a los datos (deben incluir metadatos) necesarios para validar los resultados en publicaciones científicas, los datos curados, crudos o en bruto y que especifique su DMP. En cuanto a los costes de gestión de datos, son elegibles para reembolso durante la duración del proyecto y pueden reclamarse en las condiciones definidas en el acuerdo de subvención.

Queda claro que *Horizonte 2020* exige el Acceso Abierto de todas las publicaciones científicas y desde enero de 2017 también pide la publicación en abierto de los datos de investigación con la finalidad de llegar a la comunidad científica y el público en general, buscando obtener un mayor impacto y evitar la duplicidad de esfuerzos y preservar datos para futuras investigaciones. En Europa, el acceso abierto a los datos de investigación ha sido, hasta ahora, solo un piloto para nueve áreas de proyectos financiados en el marco de *Horizon 2020*, invitando a otras áreas y programas a participar voluntariamente.

⁵⁷ Muestra la creación de un plan de gestión de datos así como manejar los datos de investigación durante y después de un proyecto de investigación también identifica las acciones y estrategias clave para garantizar que los datos de investigación sean de alta calidad, seguros, sostenibles, accesibles y reutilizables. Disponible en <https://www.openaire.eu/what-is-a-data-management-plan-and-how-do-i-create-one>.

2.1. Los datos científicos de investigación

La ciencia se está moviendo hacia una mayor apertura. Los investigadores están aprendiendo a navegar en el campo de la ciencia abierta, por lo que se están enfrentando a nuevos modelos de investigación. Y es uno de los pilares de la ciencia abierta: la apertura de los datos de investigación se está volviendo una demanda de la sociedad para los gobiernos y financiadores.

El *data sharing* consiste en compartir los datos finales de investigación entre los científicos, con el objetivo de maximizar esfuerzos y recursos. Por lo que existen numerosas razones para compartir estos datos de investigación. Algunas de ellas consisten en promover la innovación y la reutilización para nuevos usos, así como facilitar la colaboración entre usuarios de datos, creadores de datos y reutilizadores.

El permitir que otros reutilicen sus datos significa que se debe planificar desde qué punto se compartirán, ya sea desde el inicio de su proyecto de investigación o se puedan poner a disposición, en el caso, si le pertenecen, por ejemplo, a un Servicio Nacional (una entidad de Federal). Se deberá pensar en cómo se pueden compartir los datos de la investigación y qué podría limitar o prohibir el intercambio de datos.

Se comentaba anteriormente que maximiza la transparencia y la fiabilidad de los datos, favoreciendo la verificación de los resultados de investigación y aumentando el impacto y la visibilidad de la misma. Esto da pie a promover proyectos de investigación de los que provienen datos nuevos y de algunas publicaciones, lo que reduce en gran medida los costos o inversiones, evitando la duplicación de datos, como en el caso del FAIR.

El *ecosistema de datos FAIR* está compuesto por las políticas que regulan y definen los datos, los investigadores que los producen o utilizan, los planes de gestión de datos, los identificadores, los estándares, los Repositorios de Confianza y los servicios en la nube donde se almacenan los datos. A la vez, está constituido por componentes que deben desarrollarse en cuatro elementos: las capacidades o competencias, métricas, el sistema de reconocimiento y la inversión para lograrlo. Para cuestiones de este trabajo, nos centraremos solo en una parte del Ecosistema, el de los objetos de datos FAIR, que se compone de su fichero, identificador persistente, descripción de los estándares y formatos utilizados en la obtención, representación, visualización de datos y los metadatos que describen los set que deben permitir su interpretación y reutilización.

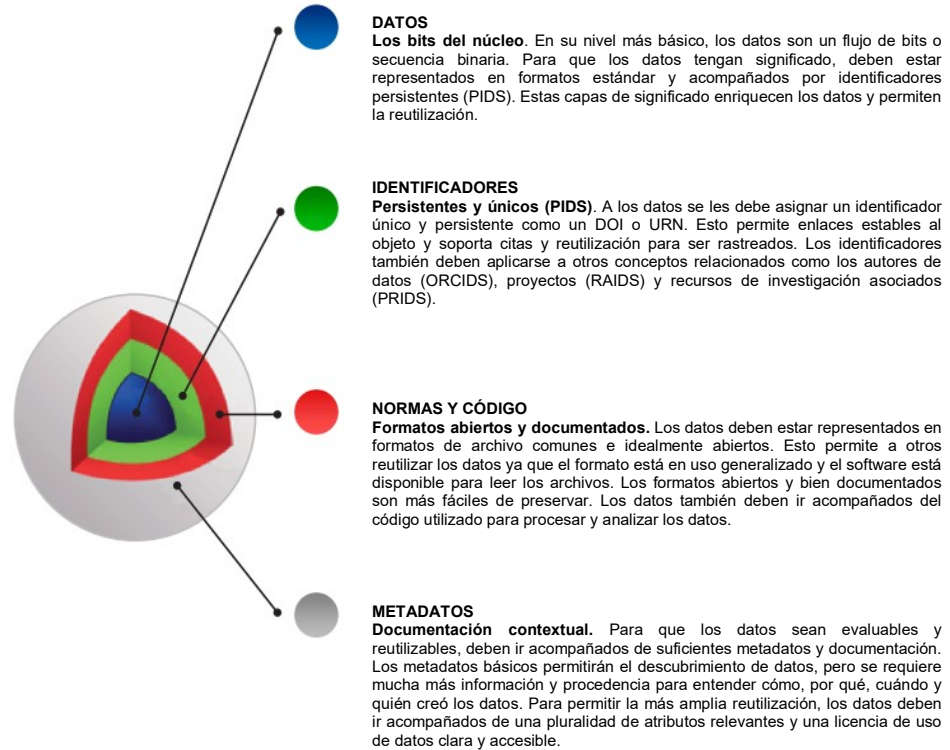


Figura 2. Modelo de datos FAIR⁵⁸.

Podemos decir que el *data sharing* se está consolidando con fuerza, sobre todo en relación con los roles y actividades que podemos desempeñar los profesionales de la información. Es por eso que en algunas bibliotecas académicas y de investigación han comenzado a desarrollar servicios para satisfacer las necesidades de gestión de datos, compartir y la preservación de los mismos.

Se presenta a continuación una taxonomía y las formas de compartir los datos de investigación:

Taxonomías de los datos de investigación	
Según su formato	Por proceso de obtención
Texto Número Imágenes	Experimentales <ul style="list-style-type: none"> - Secuencias genéticas - Cromatografías Simulaciones <ul style="list-style-type: none"> - Modelos climáticos - Modelos económicos

⁵⁸ Hodson, Simon. Hacer realidad los datos de FAIR ... y los desafíos de la interoperabilidad y la reutilización. Disponible en <http://www.eddi-conferences.eu/ocs/index.php/eddi/eddi18/paper/view/383>. (Consultado 25-06-2019).

	Observacionales - Encuestas - Experimentos irrepetibles
Según el objetivo recogida	Según la fase de investigación
Específicos - Solo de interés para un proyecto de investigación Alcance medio - De interés para una disciplina concreta De interés general - De interés para la ciencia en su conjunto e incluso de interés social	Datos preliminares - Datos recién extraídos sin ningún tipo de procesamiento. Denominados en inglés <i>raw data</i> Datos finales - Datos que ya han sido procesados y combinados con otros. Denominados en inglés <i>final research data</i> .
Medios para compartir los datos de investigación	
Comunicación formal	Comunicación informal
Repositorio de datos / bancos de datos - Centralizados - Descentralizados - Federados - Ciberestructuras	A petición - Por ejemplo, vía correo electrónico Descentralizada - Por ejemplo, a través de web personales.

Taxonomía tomada de Daniel Torres-Salinas, Nicolás Robinson-García y Álvaro Cabezas-Clavijo⁵⁹

Finalmente, podemos decir que esta forma de compartir impulsa los avances tecnológicos, acelera la investigación e involucra diferentes perfiles de especialistas dentro de una disciplina; por otro lado, la tendencia de los científicos a mostrar una actitud más abierta respecto a sus hallazgos ha ido calando progresivamente gracias a movimientos como el Acceso Abierto.

⁵⁹ Daniel Torres-Salinas, Nicolás Robinson-García y Álvaro Cabezas-Clavijo. “Compartir los datos de investigación en ciencia: introducción al datasharing”. En *El Profesional de la Informacia*. 21(2), marzo-abril 2012.

2.2.El Programa de Repositorio Institucional de Consejo Nacional de Ciencia y Tecnología de México (Conacyt)

Se han desarrollado diversos proyectos internacionales que apoyan el Acceso Abierto a la investigación. Los proyectos como *DRIVER* (Unión Europea, 2008) (Digital Repository Infrastructure Vision for European Research); *PIRUS* (Publisher and Institutional Repository Usage Statistics) para unificar estadísticas; *EuroCRIS* para unificar sistemas de gestión de la investigación y repositorios, los cuales han servido para mejorar la cooperación desde nivel local, regional, nacional e internacional, con la finalidad de ayudar y/o reforzar a las estrategias y políticas.

Pero nos enfocaremos a nivel nacional para ejemplificar las acciones tomadas por nuestro país en esta materia. En México, en el 2014, se crea una iniciativa que logra integrar el Capítulo 10 de Ciencia y Tecnología sobre Acceso Abierto a la información científica y tecnológica, creando el *Programa de Repositorios del Consejo Nacional de Ciencia y Tecnología* (Conacyt),⁶⁰ implementándose el *Repositorio Nacional* (RN),⁶¹ y a finales de ese mismo año se da pie a crear los Lineamientos Generales para el *Repositorio Nacional* (RN) y los *Repositorios Instituciones* (RI).

Esto surge tomando en cuenta que el Conacyt, desde hace 45 años, ha financiado el desarrollo tecnológico en el país y los resultados de esas investigaciones científicas que se han realizado con recursos públicos en donde todos contribuimos en cierta abierta, es decir, se busca tener acceso de manera libre y estructurada a los recursos de información. Es por eso que se implementa una estrategia de Acceso Abierto con la intención de atender las necesidades de información que se viven en este nuevo siglo y con el objetivo de cumplir las disposiciones establecidas en la legislación, asimismo con la mira de no quedar fuera del cambio mundial de ciencia abierta.

En 2015, se publican los *Lineamientos Técnicos para el Repositorio Nacional y los Repositorios Institucionales* con el fin de lograr estructurar la plataforma del RN, transformándolo en un agregador e interactuando con los RI. Desde ese momento se forma y estructura la política pública con miras a un crecimiento nacional y a una apertura de la ciencia para todos los niveles, tanto de Literatura como de Datos generados con recursos públicos. Es importante señalar que el RN utiliza estándares internacionales, buscando maximizar la diseminación de los recursos de información científica y tecnológica del país, convirtiéndose en un escaparate importante para toda la producción científica tanto de las Publicaciones Académicas que se incluyan así como de Datos que producen las Entidades que están inmersas en el Proyecto Nacional.

⁶⁰ Política de Ciencia Abierta encuentra su elemento central en el *Programa de Repositorios*, el cual se desagrega en dos componentes: Repositorio Nacional y Repositorios Institucionales.

⁶¹ El Repositorio Nacional cosecha información de diferentes repositorios institucionales que puede ser consultada desde el siguiente sitio: <http://www.repositorionacionalcti.mx/>. Actualmente el Repositorio Nacional interopera con los repositorios institucionales de los Centros Públicos de Investigación Conacyt.

Este Programa funciona a través del agregador con información de los RI, mediante la financiación de las convocatorias emitidas por Conacyt, dirigidas a instituciones que realizan investigación científica y tecnológica orientada a la creación desarrollo y/o mejora de los RI, logrando que las instituciones sigan produciendo ciencia.

Para 2016, se publican los *Lineamientos Técnicos* y de esta forma se diseña la plataforma del RN agregador de los diversos contenidos –Literatura y Datos– de los RI, con estándares internacionales, buscando maximizar la diseminación de los recursos de la información científica y tecnológica, generada en el país por las Instituciones de Educación Superior (IES).

Este programa se estructura en las siguientes etapas:

- a) En 2015, se emite la primera convocatoria dirigida a crear los RI de los centros públicos de investigación y se lanzan 27 Repositorios Interoperables con el RN, formando así el primer grupo de repositorios.
- b) La siguiente etapa consistió en el financiamiento de los RI a instituciones de Educación Pública.
- c) Finalmente, en la última fase, se incluyeron a 2 instituciones privadas. Estas tres convocatorias publicadas se llevaron a cabo de la siguiente manera:
 - 2015: 35 Repositorios Institucionales;
 - 2016: 31 Repositorios Institucionales; y
 - 2017: 15 Repositorios en total.

A través de la plataforma del RN se ponen en Acceso Abierto publicaciones académicas que son dictaminadas y su funcionamiento es financiado por convocatorias dirigidas a instituciones, así como de los Datos que producen algunas Entidades que se dedican a la investigación científica y/o tecnológica. De esta manera, se logra que instituciones que producen ciencia interoperen de forma transparente.

Los avances del Programa

El programa ha contado con distintos avances que han beneficiado a los RI tanto de Literatura como de los Datos y a los mismos investigadores para difundir su ciencia, pero casi todos los proyectos de RI se han inclinado a la Literatura, siendo pocos los que manejan Datos. Entre estos avances tenemos:

- i. Construcción de un marco teórico y técnico con el fin de ser un referente en el país para la creación de repositorios bajo lineamientos y parámetros internacionales adaptados al marco contextual de México.
- ii. Desarrollo de plataformas e infraestructura tecnológicas de un importante número de instituciones que realizan investigación científica.
- iii. Creación de capital humano así como de especialistas en esta temática que implica creación y desarrollo de repositorios, preservación, gestión digital y diseminación de recursos de información científica.

- iv. Creación de normativas e interacción con la comunidad, desarrollo de manuales, seminarios, herramientas y trabajo en ambiente de colaboración.

Según en las propuestas beneficiadas por tipo de contenido en las Convocatorias para RI que emite Conacyt en 2015 y 2016, se desarrollan los primeros Repositorios de Datos y en 2015 fueron 8 Instituciones; 2016, solo 1; y Mixtos compuestos con Datos y Literatura, 1, en 2015; y 2 en 2016.⁶²

Todos ellos dirigidos a público en general y a comunidad científica que busca información en Internet de acceso libre, así como a estudiantes, investigadores y especialistas interesados en temas de ciencia y tecnología y a los creadores, colaboradores y contribuidores de la información científica, que son los que financian y coordinan las tareas y actividades relacionadas con la ciencia y tecnología.

El impacto de este proyecto nacional en materia de Acceso y Ciencia Abierta no sólo se limita solo a las revistas, tesis, cuadernos abiertos de laboratorio, informes, etc., sino también al uso de datos, software, hardware libre que proveen los elementos para replicar (y también modificar) los resultados de recolecciones y/o experimentaciones. En general la publicación de los datos se pueden realizar en el Repositorio y además vincular todo un proceso científico con la finalidad de hacer sólidas estas iniciativas en nuestro país y aspirar a un ecosistema de innovación donde la generación del conocimiento científico supere las barreras.

Actualmente, se cuentan con 96⁶³ Repositorios Institucionales de Literatura y algunos de Datos ambos ya Interoperables con el Repositorio Nacional se está abriendo la ciencia a que sea consultada por todos los connacionales bajo una línea de libre consulta para la generación de nuevo conocimiento.

2.3. Los recursos asignados

Para la creación, desarrollo y consolidación de los Repositorios Institucionales, los cuales ya forman parte de la Política y Estrategia Nacional de Acceso Abierto en México, su desarrollo es posible a partir de los apoyos convocados por el Conacyt, en donde se abordaron acciones de promoción y fortalecimiento de la infraestructura científica, tecnológica y de innovación.

Con los nuevos principios, estrategias, modelos y herramientas se redefine la difusión de las ciencias de las Instituciones tomando nuevas estructuras y Acceso Abierto para ponerlas al alcance de todos. Como se comentó en el apartado anterior, en 2014 se publicó en México el decreto que reformó y adicionó diversas disposiciones a la legislación nacional en ciencia y tecnología, así como también en materia de Educación.

⁶² Información tomada en el *Portal del Repositorio Nacional* en el apartado de las Convocatorias emitidas por Conacyt.

⁶³ Número total de Repositorio Nacional hasta junio de 2019. Disponible en <https://repositorionacionalcti.mx/>. (Consultado: 30-06-2019).

Con base en la Estrategia 5.2 del Objetivo 5 del Programa Institucional del Conacyt 2014-2018, se instruyó y facultó para que se diseñará e integrará una Política y una Estrategia Nacional de Acceso Abierto a la Información Científica, Tecnológica y de Innovación, que permitiera el acceso y la disponibilidad a textos completos y en formatos digitales, a educandos, educadores, académicos, investigadores, científicos, tecnólogos y población en general. Uno de los elementos importantes y prioritarios de esta Política era el desarrollo y consolidación de Repositorios Institucionales (RI), así como la creación y operación de un Repositorio Nacional (RN) para México.

Por lo que los ejes de esta nueva estructura se basan en la recopilación, preservación y gestión del contenido de la información científica y tecnológica que se produce en México con fondos públicos. Asimismo, las instituciones que realizan actividades de investigación y de innovación pueden constituir su RI, ya sea por disciplinas, área científica y/o tecnológica.

Con base en lo anterior se lanzaron convocatorias para apoyar financieramente el desarrollo de RI de Acceso Abierto en instituciones de Educación Superior, centros de investigación, así como instituciones federales y estatales tanto del sector público y privado, que realizarán investigación científica y tecnológica. Los términos de estas obedecen a las disposiciones legales derivadas de la Ley de Ciencia y Tecnología (LCyT) y sus instrumentos normativos.

Los recursos asignados a cada una de las instituciones que solicitan y compiten por estas convocatorias a nivel nacional se llevan a cabo bajo 4 vertientes: 1) Creación; 2) Implementación; 3) Desarrollo y 4) Consolidación.

La evaluación y la selección de las propuestas para la asignación de los recursos económicos son sometidas a un análisis de pertinencia, el cual está a cargo de la Subdirección de Política de Ciencia y Tecnología del Conacyt, en donde se revisa que las solicitudes cumplan con los elementos establecidos y sean validados por el Comité de Ciencia Abierta (CCA). En este trabajo participan miembros del Registro Conacyt de Evaluadores Acreditados (RCEA) para hacer las evaluaciones técnico-económicas a las propuestas, las cuales deben contar con:

- Pertinencia de las propuestas en relación con los Lineamientos Generales de Ciencia Abierta y los Lineamientos Específicos para Repositorio.
- Demanda institucional por un Repositorio e impacto y beneficio potencial para la comunidad científica y tecnológica.
- Apoyo institucional de la propuesta, así como los objetivos y cronograma de actividades.
- Diseño, congruencia interna y viabilidad técnica, factibilidad técnica del proyecto y de transferencia, asimilación y adopción de los resultados del proyecto.
- Complementariedad a los esfuerzos de Conacyt en el marco de la Política de Ciencia Abierta.

En cuanto a las características del esquema de financiamiento, se apoyan los gastos e inversiones de las propuestas aprobadas y justificadas. Con base en la recomendación del Comité de Ciencia Abierta, se establece un monto autorizado y la ejecución de la propuesta en un período máximo de seis meses.

La institución beneficiada está comprometida a promover entre la comunidad científica y académica una cultura de producción y publicación en Acceso Abierto. Esta podrá expresarse públicamente en el portal del Repositorio Institucional, cumpliendo con las metas y ejercer el recurso financiero a más tardar seis meses después de que se asigne la ministración correspondiente.

2.4. La selección del software y hardware

El desarrollo de la tecnología para la puesta en marcha de Repositorios, basada en software libre, ha marcado la evolución de la gestión y en el mantenimiento. Aunque el uso de software libre no significa que a la institución no le cueste su implementación y sostenibilidad. En la mayoría de las ocasiones, la propia institución puede hacerse cargo de ello, pero en el caso de instituciones pequeñas es habitual subcontratar estos servicios a empresas especializadas para lograr y concluir su objetivo.

La idea de dar visibilidad a dicha producción genera un sinnúmero de desafíos, tanto para las universidades como para los investigadores. La capacidad que tienen los repositorios para organizar y facilitar el acceso a la producción académica y científica, ya sea por institución educativa, como por docentes, disciplinas, etcétera, hacen que estas aplicaciones sean más recomendables, puesto que permiten incluir la revisión de los documentos por parte de otros miembros de la comunidad para seguir la calidad.

Para la elección del software con características afines para la creación del Repositorio de datos, debe analizarse el funcionamiento del mismo y una serie de condiciones que van desde el tipo de servicios que se requieren, los metadatos, la aportación de datos, tipos de contenido, entre otros elementos, según las necesidades y tipología de los Datos que se van a almacenar, procesar y difundir. Además de los puntos que se enlistan a continuación:

- Elegir el software, que incluya soluciones de *Open Source* compatibles con protocolos internacionales. El análisis se debe llevar a cabo mediante un equipo interdisciplinario integrado por los especialistas de los datos e información que se incluirá, en donde la participación de bibliotecarios e informáticos aporte su experiencia sobre cómo debería funcionar.
- El uso configurable de un estándar de los Metadatos en el software de los contenidos e incorporando metodologías que ofrezcan visibilidad a los contenidos.
- Cómo se llevará a cabo la aportación de datos y fijar los mecanismos de control de calidad.

- Se debe analizar qué tipo de contenido soporta el software que empleará en el proyecto.
- Revisar los servidores: sistemas operativos, bases de datos, mecanismos de búsqueda, etcétera.
- Considerar aspectos técnicos: funcionamiento, pasos para la implementación, costos, características y proveedores del software.

Se debe buscar el proceso y la forma de cómo se personaliza la estructura de la información, es decir, los índices de consulta, así como fácil programación de un diseño web. Un punto importante consiste en revisar y analizar cuál es el software que más se utiliza en las instituciones de investigación, ya sean públicas o privadas, y evaluar su uso y aplicación en el proyecto.

A continuación se presentan varios tipos de software para implementar un Repositorio. Dentro de los más conocidos y usados tenemos los siguientes: Archimede, Bepress, Digital Commons, CONTENTdm, CDSware, CDS Invenio, DOOR, DSpace, Eprints, Equella, iTor, Fedora, Greenstone, intraLibrary, MyCoRe, PlanetDR, Research-Output Repository Platform, ResearchOutput Repository Platform, OpenRepository, OPUS, VITAL, entre otros, según las necesidades de los objetos digitales que se requieran incluir en los repositorios.

El análisis de cada uno de ellos no es el objetivo de este trabajo; sin embargo, se realizó una investigación a nivel nacional sobre qué tipo de software era el más utilizado en diversas instituciones de nivel superior y en otras entidades académicas.

Derivado de una revisión que se llevó a cabo en 96 Repositorios Institucionales que ya interoperaban con el Nacional, se identificaron una serie de características que se encuentra descritas en el Anexo 1 de este trabajo. Algunas de ellas refieren a los metadatos, comunidades, entre otras, pero sobretodo al tipo de software más utilizado y que están en funcionamiento con el RN, a continuación se enlistan:

- Dataverse, DSpace, E-prints, GeoNetwork 3.0.4.0, Omeka, Samvera (antes Hydra) y algunos desarrollos propios.

Y el de mayor uso es DSpace, por lo que consideramos mostrar cuáles son puntos clave que deben considerarse para poder llevar a cabo una instalación exitosa:

- Instalación del Java. Provee herramientas de desarrollo para la creación de programas en la unidad de red. Se pueden tener distribuidas en varias computadoras y trabajar como una sola aplicación.
- Tomcat. Este software ejecuta el repositorio y sirve de servidor para aplicaciones web. Es recomendable para entornos de desarrollo y gestión de transacciones.
- Apache Tomcat soporta todos los servicios que manejan las páginas dinámicas generadas tras la instalación del software DSpace, así como las

aplicaciones de la web. Se debe trabajar con Apache Ant y Apache Maven y, finalmente, es el que define el puerto 8080 para la consulta del localhost.

- Es importante contar con personal de cómputo que apoye el desarrollo e instalación de las aplicaciones para tener una infraestructura bien configurada y funcione durante los procesos.
- Configuración de las variables del entorno se encuentran en el sistema operativo para que funcione adecuadamente el DSpace. Se definen las rutas, para que informen al Sistema Operativo dónde y cómo ubicar el Java, así como también localizan los ejecutables. En caso de no configurar adecuadamente, causará problemas al correr los procesos.
- La creación de la base de datos PostgreSQL soporta distintos tipos de datos relacionales de objetos de código abierto. Se puede ejecutar en sistemas operativos, incluidos Linux, UNIX (AIX, BSD, HP-UX, SGI IRIX, Mac OS X, Solaris, Tru64) y Windows. Este gestor de código abierto incorpora características y funciones avanzadas en materia de seguridad, soporte de aplicaciones, seguimiento y control, rendimiento y almacenamiento de datos especiales. Es muy importante PostgreSQL, porque ahí el usuario crea la base de datos que se manejarán en DSpace.

Para concluir con la elección del software para el Repositorio, debemos considerar que tienen muchas características en común. No debemos cuestionarnos cuál es el mejor, sino cuál es el más apropiado para los servicios que se desean ofrecer por parte de nuestra institución. Por lo tanto, elegir una plataforma de software es solo uno de entre una gran cantidad de pasos y procedimientos en la implementación del Repositorio. Por ello, deben compaginarse la necesidad de innovación con los recursos que se cuentan.

2.5. El diseño de la estructura técnica usando metadatos

Para llevar a cabo el desarrollo de un Repositorio Institucional de datos deben discutirse una serie de normativas, técnicas y elementos –como son los metadatos que definirán el registro y la especificidad de cada uno de los objetos digitales albergados en el RI–. Estos son indispensables para que los contenidos sean recuperables mediante herramientas de búsqueda en Internet.

A través de los metadatos se pueden crear datos estructurados que describan el contenido, calidad y condición y otras características de un dato y además expliquen y localicen un recurso de información, y, una vez recuperado, usarlo y administrarlo.

Al trabajar con la estructura técnica del Repositorio se debe hacer mediante estándares internacionales con el fin de poder establecer una comunicación entre las diversas plataformas con las decidamos interactuar.

Es importante dejar, claro que se pueden emplear diferentes tipos de metadatos según la naturaleza del Repositorio ya sea de Datos o Literatura. En los cuales se

deben tomar en cuenta la descripción de contenido, calidad, condiciones, entre otros. Para efectos de este trabajo se consideraron tanto los metadatos para Literatura y a mayor detalle los de Datos que se explicarán más adelante.

La tipificación de los elementos de los metadatos. Generalmente, se reconocen cinco tipos para describir el objeto digital:

- Administrativos: ofrecen información sobre cómo se crearon y almacenaron los archivos que conforman el objeto digital, derechos de propiedad intelectual sobre el objeto original.
- Descriptivos: son aquellos que dependen del propio documento y sirven para representar o identificar los objetos de información digital en su fase de organización.
- De preservación: ofrecen información sobre cómo se crearon y almacenaron los archivos que conforman el objeto digital, derechos de propiedad intelectual, resolución de formato y software, justamente con los datos necesarios para llevar a cabo el proceso de conservación.
- Técnicos: se refieren a los elementos creados por o para un sistema automatizado, y referirán al funcionamiento del sistema de documentos y datos sobre los tiempos de respuesta.
- De uso: generalmente, son creados de forma automática, relativos al nivel de utilización y al tipo de usuario de un determinado servicio.

Debemos asegurarnos que los metadatos mantendrán su capacidad de interpretación y su utilidad para que los registros puedan almacenarse y conservarse a lo largo del tiempo.

Funciones de los metadatos

- Permiten organizar mejor los recursos electrónicos y facilitan una identificación digital.
- Describen recursos, identifican, exhiben a otros recursos o similares para ubicarlos.
- Facilitan la organización de los recursos a medida que la web, los sitios o los portales crecen; se organizan enlaces basados en la audiencia.
- Interoperan la posibilidad de cambiar datos entre hardware en diferentes plataformas operativas, así como interfaces mediante estructura de datos.
- Identificación digital: varios esquemas de metadatos contienen elementos para cargar números normalizados que permitan identificar unívocamente la obra o el recurso mediante el URL, PURL o un DOI.
- Archivo y preservación. Por lo regular, es probable que muchos recursos digitales no puedan ser usados en el futuro. La información electrónica puede ser alterada por el cambio de la tecnología del hardware o el software. La migración de formatos y la emulación de programas son estrategias para evitar los inconvenientes.
 - Los metadatos se pueden expresar en XML, lenguaje independiente del software e intercambiable entre distintos sistemas. Se expresa a

partir de una Definición de Tipos de Documentos (DTD), especificando las etiquetas que se emplearán en el documento.

Formatos de metadatos

Se presentan varios formatos destinados a diversos objetos digitales así como a publicaciones académicas, esto con la finalidad de ofrecer un panorama sobre este tipo de materiales y que se enlistan a continuación:

- **Dublin.** Es una infraestructura operacional, que se compone en elementos y generalmente usa XML.
- **Mods.** Contiene 20 elementos principales, sub-elementos y atributos que pueden ser usados en uno o más elementos. Está pensado para ser utilizado por especialistas del ámbito de las Ciencias de la Información y es fácil de exportar a otros esquemas.
- **Mets.** Permite la codificación de los diferentes tipos de metadatos y sobre todo de la preservación y recuperación de los objetos. Se componen en 7 secciones.
- **Mix.** Se centra específicamente en información técnica sobre la creación y modificación de imágenes compatible con estándares como METS, al formar parte de un servicio de preservación digital.
- **Premis.** Por lo general, cuando alguien se refiere a *Premis* se refiere al Diccionario de Datos. A veces pueden referirse al esquema XML, al grupo de trabajo, o a todo, incluida la actividad de mantenimiento. Podemos decir que es un conjunto de unidades semánticas fundamentales que deben entender los repositorios para llevar a cabo sus funciones de preservación.
- **LOM.** Se centran en el conjunto mínimo de propiedades que permiten que los objetos educativos sean gestionados, ubicados y evaluados y es exclusivamente metadatos para objetos educativos.
- **OME-XML, MIBBI, AgMES** (Agricultural Metadata Element Set). Todos estos utilizados en Biología.

Para datos en general y de investigación tenemos DataCite –el cual se explicará más adelante–, CERIF, DCAT Data Catalog Vocabulary; en el área multidisciplinar, EML, ISO 19115, OAI-ORE (Open Archives Initiative Object Reuse).

Los metadatos en el área de las Ciencias de la Tierra son:

- ISO 19139 e ISO 19139 GML 3.2. Permiten ver y editar un documento de metadatos completo que cumple el estándar de la información, así como también archivos exportados utilizan el espacio de nombre GML 3.2 y, por consiguiente, pueden validarse con las versiones de los Esquemas XML geográfica, AVM Astronomy Visualisation Metadata, CIM, Common Information Model, entre otros.
- Metadatos FGDC CSDGM. Este estilo permite ver y editar metadatos de acuerdo con las directrices Estándar de Contenidos Geoespaciales Digitales (CSDGM) de CSDGM, exportar metadatos en el formato XML de esta norma y validarlos con CSDGM XML DTD.

Aplicación de los metadatos

Como se mencionó anteriormente la aplicación de estos puede ser en recursos, objetos digitales como videos, imágenes, mapas, recursos educativos, además de recursos geoespaciales, datos geográficos entre otros más. A continuación algunos ejemplos:

- Metadatos para recursos bibliográficos: comprenden el título del recurso, el individuo u organización responsable de su creación, el lugar donde fue creado, la descripción física, etc.
- Metadatos para recursos geoespaciales: un ejemplo es la Norma de Contenidos para Metadatos y Geoespaciales Digitales, establecida por el Comité Federal de Datos Geográficos.
- Metadatos para recursos de redes: describen los recursos conectados en la red. Incluye elementos bibliográficos como título, creador, editorial, etc., y también por el tipo y formato del recurso, relaciones entre varios recursos y los derechos de propiedad intelectual.
- Metadatos para recursos educativos: se requieren elementos específicos, además de los metadatos bibliográficos comunes. Por ejemplo: mientras que un recurso educativo tiene un autor, lugar y fecha de publicación, también se puede haber creado para un público educativo particular, que va desde pre-escolares hasta universitarios. Para capturar estos elementos, se han desarrollado diferentes estándares.

2.5.1. Dublin Core

El Dublin Core Metadata Initiative (DCMI) es una organización que apoya la innovación en el diseño de metadatos y las mejores prácticas en la ecología de los metadatos. Tiene sus raíces en Chicago, en la 2^{da} Conferencia Internacional de la World Wide Web, en octubre de 1994. Yuri Rubinsky de SoftQuad, (quien presidió los paneles sobre el futuro de las herramientas de autoría de HTML y Web) junto con Stuart Weibel y Eric Miller de OCLC, presentaron trabajos sobre publicaciones académicas en la Web y sobre los principales problemas en la prestación de servicios bibliotecarios basados en la Web y la dificultad de encontrar recursos. En ese taller se analizó cómo los recursos facilitarían la búsqueda y recuperación, además de poder clasificarla. El resultado de esos análisis fueron llamados *Metadatos del núcleo de Dublín*.

En 2001, el formato del Taller se amplió para incluir tutoriales, documentos y pósteres de conferencias, revisados por pares, ofreciendo a la comunidad de metadatos una mayor oportunidad de aprendizaje, intercambio de ideas y desarrollo de estándares de metadatos de Dublin Core (DC).

Este modelo de metadatos está encaminado a fomentar la adopción extensa de los estándares interoperables y a promover el desarrollo de los vocabularios especializados para describir recursos y permitir sistemas más inteligentes del descubrimiento del recurso. Las implementaciones de Dublin Core usan

generalmente XML y se basan en el Resource Description Framework (RDF) y se define por ISO, en su norma ISO 15836 –en 2009–, y la norma NISO Z39.85-2012. Desde 2003, Dublin Core adquiere la categoría de Norma ISO 15836. A la vez, este es un sistema de 15 definiciones semánticas descriptivas o elementos, mismas que pueden ser optativas y repetibles y se ha convertido en un estándar de metadatos que intenta describir un gran número de recursos Web.

Los elementos poseen etiquetas descriptivas que pretenden transmitir un significado semántico a los mismos. Cada elemento es opcional y puede repetirse. Además, los elementos pueden aparecer en cualquier orden. Aunque algunos entornos, como HTML, no diferencian entre mayúsculas y minúsculas. Es recomendable escribir correctamente cada metadato, según su definición, para evitar conflictos con otros entornos, como SGML y XML.



Figura 3. Los elementos Dublin Core se pueden clasificar en tres grupos que indican la clase o el ámbito de la información (Elaboración propia)

Su estructura se basa en elementos básicos y calificadores, utiliza etiquetas basadas en palabras. Todos los elementos son opcionales y repetibles. Es un estándar de estructura, que no está regido por la reglas de contenido, sino que se sugiere la normativa a aplicar. Se codifica en HTML, XML y RDF, formatos más divulgados y aceptados para representación y recuperación de información en la web.

Además, para la descripción de recursos web es concebido el *DLO* (Unidad Documental o al Documento Digital Mínimo). Forma parte de una colección digital, al cual se le aplican metadatos para su descripción y recuperación.

Es uno de los formatos más divulgados y aceptados para representación y recuperación de información en la web. Es el que más afluencia ha ejercido en lo que respecta al desarrollo general del concepto y teoría de los metadatos.

Dentro de sus fortalezas, Dublin Core es simple de utilizar y aplicar en los recursos. Posee independencia sintáctica así como de alto nivel de normalización formal, lo que lo convierte en un estándar. Su infraestructura es operacional y se

lleva de la mano con el desarrollo de la Web Semántica y es el más utilizado sólo en los Repositorios que incluyen Literatura. Su esquema de metainformación es el más empleado a nivel mundial.

2.5.2. DataCite y OpenAIRE

Con base en el desarrollo del Repositorio de datos y objetivo central de este trabajo, se describirán ambos esquemas utilizados e importantes para la recopilación e importación de metadatos. El primero, DataCite, es una organización global sin fines de lucro que proporciona identificadores como el DOI, además de asignación de los metadatos y servicios que permiten a la comunidad de investigación identificar, ubicar, acceder, conectarse y citar datos de investigación con la finalidad de que sean visibles y accesibles. También proporciona apoyo a los investigadores, centros de datos, editores de revistas y agencias de financiamiento mediante una colaboración con una red global.

Asimismo, desarrolla servicios para mejorar la experiencia de gestión de los identificadores mediante la conexión e intercambio en ecosistemas de investigación que se pueden consultar directamente para encontrar los datos, obtener estadísticas, explorar conexiones de libre acceso y revisión en una interfaz de búsqueda integrada también filtra y extrae todos los detalles de una colección de millones de registros

Las organizaciones de investigación se unen a DataCite para poder asignar identificadores a sus resultados de investigación. De esta manera, promueven el intercambio de datos y citas a través de esfuerzos de construcción de la comunidad y actividades de divulgación. Permite, además, el archivado de datos, lo que facilita que los resultados sean verificados y reutilizados para nuevos estudios en el futuro.

El servicio de protocolo que ofrece es mediante DataCite o OAI-PMH, exponiendo los metadatos almacenados en el DataCite Metadata Store (MDS), iniciativa de archivos abiertos para la recolección de metadatos y facilita un conjunto de servicios que permite la exposición y la recolección de metadatos del repositorio. El servicio es abierto para todos y puede ser accedido por los recolectores y aplicaciones (API) que cumplan con OAI-PMH, mediante su base. Disponible en <https://oai.datacite.org/oai>.

Este protocolo consta de seis verbos que especifican el tipo de servicio que se ofrece:

1. *Identify*: se utiliza para recuperar información sobre el Repositorio.
2. *ListIdentifiers*: recupera encabezados de registros desde el Repositorio.
3. *ListRecords*: recopila registros completos del Repositorio.
4. *ListSets*: recupera la estructura del conjunto del Repositorio.
5. *ListMetadataFormats*: enumera los formatos de metadatos disponibles que el Repositorio puede difundir.

6. *GetRecord*: se utiliza para recuperar un registro individual del Repositorio.

La recolección selectiva se puede realizar mediante el uso de los parámetros que lo acompañan. Algunos de estos son:

- *Identifier*: especifica un identificador de registro específico.
- *MetadataPrefix*: especifica el formato de metadatos en el que se devolverán los registros.
- *Set*: especifica el conjunto al que deben pertenecer los registros devueltos.
- *From*: especifica que los registros devueltos deben haber sido creados/actualizados/eliminados en o después de esta fecha.
- *Until*: especifica que los registros devueltos deben haber sido creados/actualizados/eliminados en esta fecha o antes.
- *ResumptionToken*: un token es proporcionado previamente por el servidor para reanudar una solicitud donde se detuvo por última vez.

Estos verbos y los parámetros se pueden combinar para emitir solicitudes al servicio, tales como:

- <https://oai.datacite.org/oai?verb=Identify>
- https://oai.datacite.org/oai?verb=ListIdentifiers&metadataPrefix=oai_dc
- https://oai.datacite.org/oai?verb=ListRecords&from=2011-06-01T00:00:00Z&metadataPrefix=oai_dc

Formatos de metadatos disponibles en DataCite OAI-PMH:

- **OAI Dublin Core (oai_dc)**, como requisito mínimo para el cumplimiento de OAI-PMH, los metadatos deben estar disponibles en el formato Dublin Core de OAI.
- **OAI DataCite (oai_datacite)**. Este formato de metadatos se ha establecido específicamente para la difusión de registros de DataCite, utilizando OAI-PMH. Además de los metadatos originales, contiene varios otros elementos que describen la versión de los metadatos y el centro de datos de registro.
- **DataCite Direct (datacite)**. Este formato de metadatos contiene solo los metadatos DataCite originales sin adiciones ni alteraciones. Debido a que existen múltiples versiones de los metadatos. No hay un solo esquema al que todos se adherirán. No es compatible con la versión 2.0 de OAI-PMH.

Esquema de metadatos DataCite

Sus principales propiedades son un conjunto de elementos individuales que se usan para describir los datos. A cada metadato-elemento en el esquema se le asigna un nombre y un significado además de ser jerárquico, flexible, extensible y utilizado por OpenAIRE para recolectar metadatos de repositorios.

Mandatory	Recommended	Optional
Identifier	Subject	Language
Creator	Contributor	Alternate ID
Title	Date	Size
Publisher	Resource Type	Format
Publication year	Related identifier	Version
	Description	Rights
	GeoLocation	

Figura 4. Niveles obligatorios para las propiedades de metadatos DataCite.⁶⁴

Los cuatro niveles de obligación para las propiedades de metadatos:

- Obligatorio (M): deben ser proporcionados; es decir, el campo siempre debe estar presente en el registro de metadatos. No se permite un elemento vacío.
- Obligatorio cuando sea aplicable (MA): cuando el valor del campo se puede obtener, debe estar presente en el registro de metadatos.
- Recomendadas (R): son opcionales, pero se recomienda el uso del campo.
- Opcionales (O): la propiedad puede usarse para proporcionar información complementaria sobre el recurso.

Se solicitan los niveles para que los usuarios ingresen información necesaria de las propiedades al crear un registro, y mejora la calidad de los servicios así como la recuperación.

Aquellos centros de datos que busquen mejorar las perspectivas de sus metadatos se encontrarán, citarán y vincularán a la investigación original. Es imprescindible aportar el Conjunto de Propiedades Recomendado y Obligatorio para que los registros sean sustentados correctamente, con sus datos que lo respalda. El Conjunto de Propiedades Obligatorio y Recomendado así como sus subpropiedades son especialmente valiosos para los buscadores de información y los proveedores de servicios agregados, como los indexadores. DataCite recomienda encarecidamente la inclusión de los metadatos identificados como Recomendados, a fin de lograr una mayor exposición para el registro de metadatos del recurso y, por lo tanto, la investigación subyacente en sí.

Esquema de metadatos en DataCite cambios que incluyen cada una de las versiones			
Borrador del 20 de marzo de 2019. Los cambios en esta versión incluyen	23 de octubre de 2017	19 de septiembre de 2016	16 de octubre de 2014
Versión 4.2	Versión 4.1	Versión 4	Versión 3.1.
- Adición del nuevo tipo de fecha "Withdrawn" - Adición de un nuevo	- Permitiendo múltiples polígonos por <i>GeoLocation</i>	- Cambiando <i>resourceTypeGeneral</i> de opcional a	- Nuevo atributo de afiliación para <i>Creator</i> y <i>Contributor</i>

⁶⁴ "The DataCite Metadata Schema. What is the DataCite metadata schema?". Disponible en <https://project-thor.readme.io/docs/datacite-metadata-schema>. (Consultado: 25-06-2019).

<p>par <i>relationType</i> pair: <i>IsObsoleteBy</i> and <i>Obsoletes</i></p> <ul style="list-style-type: none"> - Adición del nuevo tipo de identificador relacionado <i>relatedIdentifierType</i> "w3id" - Adición de nuevas subpropiedades de derechos: identificador de derechos, <i>rightsIdentifierScheme</i> - esquemaURI - Adición del atributo de lenguaje XML a las propiedades <i>Creator</i>, <i>Contributor</i> y <i>Publisher</i> para nombres de organizaciones. 	<ul style="list-style-type: none"> - Adición de nuevas subpropiedades opcionales "inPolygonPoint" para el polígono - Adición de un nuevo tipo <i>dateType</i> "Other" - Adición de un nuevo tipo de recurso <i>resourceType</i> "DataPaper" - Adición de tres nuevos pares: <i>IsDescribedBy</i> and <i>Describes</i>, <i>HasVersion</i> and <i>IsVersionOf</i>, <i>IsRequiredBy</i> and <i>Requires</i> - Adición de una nueva subpropiedad para <i>Date</i>: <i>dateInformation</i> - Adición de un nuevo atributo opcional para <i>creatorName</i> y <i>ContributorName</i>: <i>nameType</i>. Controlled list: personal, organizational - Adición de un nuevo atributo opcional "resourceTypeGeneral" attribute for <i>relatedIdentifier</i>. Controlled list is identical to existing <i>resourceTypeGeneral</i> attribute - Adición del atributo <i>lang</i> opcional a la <i>Rights</i> property 	<p>obligatorio</p> <ul style="list-style-type: none"> - Adición de una nueva propiedad: <i>FundingReference</i>, con subpropiedades <i>funderName</i>, <i>funderIdentifier</i>, <i>awardNumber</i>, <i>awardURI</i> and <i>awardTitle</i>. <i>Deprecation of contributorType</i> "funder" - Adición de nuevas subpropiedades opcionales para <i>creatorName</i> y <i>contributorName</i>: <i>familyName</i> y <i>givenName</i> - Adición de una nueva opción relacionada con el identificador de identificador <i>relatedIdentifierType</i> option "IGSN" - Adición de una nueva subpropiedad <i>GeoLocation</i> "geoLocationPolygon", y cambio de la definición de las subpropiedades de <i>GeoLocation</i> existentes <i>geoLocationPoint</i> y <i>geoLocationBox</i> 	<ul style="list-style-type: none"> - Nuevos pares <i>IsReviewedBy/Reviews</i> and <i>IsDerivedFrom/IsSourceOf</i> - Nuevo <i>contributorType</i>: <i>DataCurator</i> - Nuevos tipos de identificadores relacionados: arXiv y bibcode
---	--	---	---

Existe un Grupo de trabajo de metadatos que busca hacer que los datos sean citables y accesibles; es decir, equipar los conjuntos de datos con metadatos (descripciones y datos y cifras sobre los datos) que cumplan con los estándares básicos y se adhieran a un esquema uniforme y consistente (Elaboración propia).

OpenAIRE

El segundo esquema de metadatos es un proyecto europeo de apoyo a Open Science. Por otro lado, es una red que cuenta con más de 50 instituciones, que trabajan para implementar políticas efectivas de Open Access y Open Science. Mediante la red de 34 países de la Unión Europea (UE) llamada *OpenAIRE's National Open Access Desks (NOADs)*, recopilan y actualizan información sobre la ciencia abierta a través del entrenamiento de política-infraestructura de forma periódica que garantice contenido actualizado.

Está a la vanguardia de los movimientos de Acceso Abierto y de datos abiertos en Europa. También es el encargado de respaldar la política de datos abiertos al

proporcionar un repositorio pionero en código abierto, Acceso Abierto y datos abiertos siendo este Zenodo, creado por OpenAIRE y CERN.

Se lanzó en 2013, permitiendo a los investigadores en cualquier área temática cargar archivos de hasta 50 GB en su interior. Esta herramienta de investigación abierta proporciona lugar para que los investigadores depositen publicaciones, conjuntos de datos y otros artefactos de investigación como códigos, carteles, presentaciones. Este Repositorio no impone requisitos de formato, tamaño, restricciones de acceso o licencia. No está restringido a un financiador o una nación.

También involucra a los interesados para una implementación efectiva de Open Science mediante el diálogo científico abierto para las políticas y su implementación en Europa; además, proporciona servicios científicos abiertos a través de la interoperabilidad, conectando la investigación y permitiendo a investigadores, proveedores de contenido, financiadores y administradores de investigación adoptar fácilmente la ciencia abierta.

Basados en estándares comunes globales para vincular la investigación (publicaciones, datos, software) a sus creadores (investigadores, instituciones, financiadores) con la finalidad de que haya transparencia, reproducibilidad y la garantía de calidad de la investigación, permitiendo informes, monitoreo y análisis que pueden usarse para saber el movimiento de nuestra ciencia.

Emplea iniciativas para promover y compartir políticas y protocolos de acceso comunes para todos los resultados de la investigación. Como resultado, se promueven estándares de datos comunes e intercambio de contenidos para la transición hacia un ecosistema de investigación global eficaz y abierto. Esto abre canales de comunicación, así como nuevos horizontes para la investigación, poniendo a disposición toda la información de investigación de forma gratuita para que los proveedores externos creen servicios de valor agregado y se fomente una mejor comprensión del conocimiento científico.

Las directrices de OpenAIRE

Estas apoyan a los administradores de los Repositorios a difundir publicaciones, conjuntos de datos, sistemas CRIS y Repositorios de otros productos de investigación, respectivamente, a través del protocolo OAI-PMH, para integrarse con la infraestructura de OpenAIRE. Las diferentes directrices se enlistan a continuación:

1. **Repositorio de literatura.** La versión más reciente de las Directrices de OpenAIRE para administradores de este tipo de Repositorios es la versión 4 y presenta los siguientes cambios y elementos:
 - i. Perfil de aplicación y un esquema basado en Dublin Core y DataCite que incluye un nuevo OAI-metadataPrefix.

- ii. Soporte de esquemas de identificación para autores, organizaciones, financiadores, recursos académicos.
- iii. Introducción de Vocabularios Controlados COAR.
- iv. Cumplimiento de la Política de Adquisición de Contenido de OpenAIRE, publicada el 5 de octubre de 2018.

2. Repositorio para datos. El contenido del Repositorio debe ser configurado bajo el setSpec *openaire_data*. El recolector usará el valor setSpec para realizar la recolección selectiva y determinará su comunicación, por lo que recolectará los conjuntos de datos, cumpliendo al menos uno de los siguientes criterios:

- El conjunto de datos es el resultado de un proyecto(s) de investigación ya sea financiado o no.
- Los datos deberán vincularse a una publicación en el espacio de información de OpenAIRE.
- Los dos anteriores puntos se deben o pueden determinar por la infraestructura de OpenAIRE.

Estas directrices pueden emplearse en un conjunto de datos de financiamiento o a un enlace de una publicación relacionada con el Repositorio de Datos. En el caso de una recopilación posterior a un Repositorio de Literatura con enlaces al conjunto de datos, el conjunto de datos se expondrá en el portal de OpenAIRE, pero variará el período de recolección.

Para que las propiedades del conjunto de datos sean compatibles con la infraestructura de OpenAIRE, se adoptó esquema de metadatos DataCite v3.1 con algunos ajustes en cinco puntos que a continuación se enlistan:

a) Derechos de acceso e información de licencia

- i. Aplica un esquema de codificación en la propiedad DataCite, utiliza los derechos de acceso para permitir una mejor experiencia de usuario al declarar los derechos de acceso claros y explícitos. Los derechos de acceso se especifican utilizando la propiedad *16. Rights (MA)*. Además, se recomienda incluir información disponible sobre la licencia.

```
1 <rightsList>
2 <rights rightsURI = "info: eu-repo / semantics / openAccess" />
3 </rightsList>
```

```
1 <rightsList>
2 <rights rightsURI = "info: eu-repo / semantics / openAccess" />
3 <rights rightsURI = "http://creativecommons.org/licenses/by/4.0/" >
4   Atribución de Creative Commons 4.0 Internacional
5 </rights>
6 </rightsList>
```

b) Información del financiamiento

- i. Es necesario el uso específico de la propiedad para identificar los enlaces a fuentes de financiamiento. Uno de sus objetivos es vincularlo a los resultados de la investigación. Por lo que se debe usar la propiedad para relacionar al conjunto de datos con la información financiada. [Consultar los apartados: 7. Contributor (MA/O), 7.1 contributorType (MA/O), 7.2 contributorName (MA/O), 7.3 nameIdentifier (MA/O) and 7.3.1 nameIdentifierScheme (MA/O)].
- ii. Esta propiedad del contribuyente permite la identificación única del financiador, sin importar que lo haya hecho total o parcialmente al conjunto de datos que se describan.

c) Publicaciones relacionadas e información de conjuntos de datos

- i. Se pretende obtener detalles específicos de qué conjuntos de datos se exponen. Es decir, los conjuntos de datos relacionados con la publicación deben ser definidos explícitamente, con el fin de vincularlos. Por lo que se visualizarán de 1-2 días después de la recolección (un Repositorio se cosecha una vez a la semana, en promedio). Si la infraestructura deduce automáticamente el enlace, puede tardar hasta un mes después de la recolección antes de que el conjunto de datos se exponga. Por lo tanto, es obligatorio, cuando corresponda, proporcionar enlaces a publicaciones y conjuntos de datos cuando estos enlaces estén disponibles en el Repositorio, y así garantizar una exposición más rápida del conjunto de datos.
- ii. Por lo que se recomienda exportar enlaces a publicaciones y conjuntos de datos relacionados, es decir, las propiedades y los atributos son obligatorios cuando corresponda, citado a detalle en el apartado 12. *RelatedIdentifier (MA)*.

d) Identificadores relacionados

- i. El esquema de metadatos de DataCite permite vincular publicaciones y conjuntos de datos mediante el uso de identificadores persistentes para identificar de forma única el recurso que se describe como (A), generalmente un conjunto de datos, pero no limitado a eso, y el recurso relacionado (B) en el caso de una publicación o datos. Por lo que acepta otros esquemas de identificadores persistentes y no solo un DOI, tales como: ARK, Handle, PURL, URN, citados en el apartado 1.1 *identifierType (M)*.

e) Información de fecha de publicación

- i. Para estas directrices son relevantes y obligatorias dos tipos de fechas. Cuando los datos se ponen a disposición, se publican con una fecha y otro caso es cuando se hayan

cargado a una base de datos, siendo esta la fecha en que los datos fueron “emitidos”.

- ii. En otras ocasiones, los datos pueden ser embargados por un período. Este tipo de información debe ser administrada por el proveedor de datos y expresada mediante la exportación de una fecha “disponible” para indicar el fin de un período de bloqueo y una fecha “aceptada” para indicar el inicio de un período. Las propiedades opcionales de 8. DataCite (M) y 17. Description (MA) son obligatorias en OpenAIRE.

3. Gerentes de CRIS

Está dirigido a la orientación de los administradores para que expongan sus metadatos de manera que sean compatibles con la infraestructura de OpenAIRE. Al implementar las Directrices, los administradores apoyan la inclusión y, por lo tanto, la reutilización de metadatos en sus sistemas dentro de la infraestructura.

El Formato de Información de Investigación Común Europea, (*Common European Research Information Format* –CERIF-CRIS, por sus siglas en inglés–), es un modelo de datos estándar para información de investigación y recomendado por la Unión Europea a los Estados miembros, dedicado a la interoperabilidad de los sistemas de información de investigación y modelo de dominio y de datos formal. Incluye un mecanismo para construir perfiles XML (subconjuntos especializados) para escenarios específicos de intercambio de información, administrando la recolección e importación de metadatos de los sistemas CRIS (Disponible en <http://www.eurocris.org>). Existen 3 versiones, pero se está considerando la versión 1.1.1., de diciembre de 2018, doi: 10.5281 / zenodo.2316420.

4. Gestores de Repositorios de Software

Las Directrices para los Gestores de Repositorios de Software 1.0 proporcionan indicaciones sobre cómo hacer que los productos de software sean citables para que los usuarios identifiquen el ecosistema de comunicación académica, lo que favorecerá la exposición, la visibilidad y la reutilización del contenido del Repositorio, aumentando significativamente la consulta.

También ofrecen visibilidad inmediata al software como un “producto de investigación citable”. El software de investigación está actualmente disponible en los siguientes tipos de Repositorios de Comunicación Académica:

- Repositorios institucionales: las descripciones de software se proporcionan como registros de metadatos de Dublin Core.
- Repositorios de datos: las descripciones de software se proveen como registros de metadatos DataCite/DataVerse. Los metadatos están orientados a la reutilización del software, en lugar de citas.
- Repositorios de software: la mayoría de estos Repositorios de Software están desarrollando a la GitHub, desarrolladores que trabajan juntos para alojar y revisar el código, administrar proyectos y construir software. De

esta manera, los metadatos están orientados a la reutilización del software, en lugar de citas; en algunos casos, se definen como Repositorios de Investigación, por lo que incluyen metadatos para el descubrimiento y la cita.

5. Otros Productos de Investigación (ORP) Administradores de Repositorios

Al implementar estas pautas, basadas en DataCite y en las de software OpenMinTed SHARE-OMTD,⁶⁵ se utilizan esquemas de metadatos para los recursos de software. Por lo que los administradores de Repositorios les permitirán a los autores incorporar su producto de investigación en la infraestructura de OpenAIRE para que puedan descubrir y utilizar los servicios de valor agregado.

2.5.3. GeoNetwork

Dentro del ámbito de los Repositorios de los datos en ciencias de la tierra, es uno de los software que se emplea para administrar los contenidos y que ha aumentado su implementación en el área de Ciencias de la Tierra es *GeoNetwork*⁶⁶, ésta aplicación administra recursos de referencia espacial, edición de metadatos, así como un visor de mapas web interactivo.

Este comenzó como un Sistema de Catálogo de Datos Espaciales para la Organización de las Naciones Unidas para la Agricultura y la Alimentación (FAO), el Programa Mundial de Alimentos de las Naciones Unidas (PMA) y el Programa Ambiental de las Naciones Unidas (PNUMA) utilizando como base las Infraestructuras de Datos Espaciales en todo el mundo y forma parte de la Open Source Geospatial Foundation (OSGeo), siendo un proyecto financiado que proporciona servicios de gestión de información espacial. Diseñado para organizar y facilitar el acceso a recursos de cartografía, bases de datos espaciales y metadatos asociados a través de un único punto de entrada, evitando duplicidad de información y fomentando su intercambio. Se basa en estándares abiertos (OGC) y permitiendo trabajar con datos descentralizados procedentes de varios Repositorios.

Cuenta con una infraestructura de gestión de la información geográfica en la cual analiza los datos y puede añadir el registro de metadatos utilizando la herramienta de inserción de metadatos XML. Los estándares de metadatos que maneja esta fuente abierta son la ISO 19115: 2003, aprobada por la comunidad internacional en abril de 2003 como una herramienta para definir metadatos en el campo de la información geográfica, y el FGDC, el estándar de metadatos adoptado en los

⁶⁵ OpenMinTeD. Open Mining Infraestructure for text & Data (OpenMINTED): Guedelines https://guidelines.openminted.eu/guidelines_for_providers_of_sw_resources/recommended_schema_for_sw_resources.html. (Consultado: 25-06-2019).

⁶⁶ GeoNetwork opensource. El proyecto GeoNetwork opensource es una aplicación informática de software libre y código abierto de catalogación para recursos referenciados al espacio geográfico. <https://geonetwork-opensource.org/manuals/3.6.x/es/index.html>. (Consultado: 28-06-2019).

Estados Unidos por el Gobierno Federal del Comité de Datos Geográficos, que admite también el estándar internacional Dublín Core para la descripción de documentos en general.

Contiene plantillas extensibles con nuevos elementos para satisfacer necesidades especiales que se requieran para describir los datos geográficos vectoriales o ráster basados en ISO19139 y en ISO19115.

El registro de los datos

Los campos dentro de esta aplicación para compilar el registro de metadatos son los siguientes:

Título, Fecha de creación o publicación, Resumen, Idioma utilizado para documentar datos, Categoría de tema, Escala, Mantenimiento y Frecuencia de actualización, Autor de metadatos, Idioma Utilizado para documentar metadatos.

Existen campos obligatorios y opcionales (si la información se encuentra disponible):

Propósito, Palabras clave, Formulario de presentación, Estado, Tipo de representación espacial, Ubicación geográfica, Información del sistema de referencia, Extensión temporal, Datos Información de calidad, Restricciones de acceso y uso, Punto de contacto, Información de distribución y Recursos en línea.

El registro de los datos tanto gráficos, de documentos, archivos pdf y cualquier otro tipo de contenido. Soporta la edición de metadatos multilingüe, la validación y sugerencias para mejorar la calidad. Y además la Geopublicación de capas para publicar geodatos en servicios OGC (por ejemplo, GeoServer).

Dado que el estándar de metadatos de mayor uso en *GeoNetwork* con la ISO19115: 2003 para la información geográfica ahora es el estándar común preferido, muchos tienen la necesidad de migrar metadatos a nuevos estándares proporcionando una funcionalidad de importación (y exportación) basadas en XSLT.

Además se recomienda el uso de esta tecnología por ser un servidor de catálogo Open Source completo y que integra tanto las funcionalidades requeridas para la introducción de metadatos (editor de metadatos) como para su explotación (búsquedas y publicación).

2.6. La generación de comunidades y subcomunidades

La estructuración de la información y contenido que se integrará al repositorio de datos, debe contar con una organización definida con base en las necesidades y particularidades de cada Institución, Entidad Colegio u Organismo es importante

considerar el objetivo del porqué se está creado una herramienta y qué alcance debe de tener.

Es necesario determinar los formatos, servicios y el papel de la biblioteca porque de ello depende el tipo de comunidades se puedan crear y diseñar.

Partiremos haciéndonos las siguientes interrogantes: ¿qué tipo de datos se aceptarán? aquellos que la institución requiera y sobre todo cuente con los derechos y permisos para publicar, se tiene que definir este procedimiento para incluir contenido de calidad.

Es decir cada servicio del Repositorio de datos organiza su contenido bajo una estructura académica algunas Universidades lo establecen según los centros o departamentos de investigación, áreas académicas, administrativas sobre todo es con base en lo que se publicará en esta herramienta. Se presentan algunas formas por la cuales pueden estructurar las comunidades necesarias o área que incluirán en el Repositorio de datos:

- *Comunidades departamentales*, constan de departamentos, centros de investigación y grupos ya existentes, con directrices de contenido bajo líneas de investigación y solo pueden depositar los pares bajo una revisión y aceptación de la comunidad.
- *Comunidades por áreas temáticas*, todo el personal académico puede aportar datos, a través de terceros. Y es revisado por el personal académico de la biblioteca antes de publicarlo o hacerlo visible en el repositorio.
- *Comunidades por tipología*, dentro de los repositorios de datos estructurarlos tiene una lógica peculiar en donde el dato representa una variable y/o una asignación de valores además se necesita considerar el uso de software y la complejidad para su recolección y posterior tratamiento.
- *Comunidades de interés* se crean bajo un perfil dependiendo de los intereses y acuerdos de sus integrantes determinando una necesidad fija a cubrir mediante una justificación.

Una vez determinadas el tipo de las comunidades, se entenderá de forma más ordenada y constituida la estructura general del Repositorio de datos. Como siguiente paso nos enfocaremos a otro nivel al interior de las Comunidades, la descripción de las subcomunidades y así sucesivamente hasta llegar a las colecciones que facilitan organizar y clasificar el contenido del Repositorio.

Para ello, se llevó a cabo una revisión del tipo de software que más se utiliza en los Repositorios Institucionales tanto de Literatura como de Datos en México, y en al menos 100 instituciones entre gubernamentales, públicas, privadas y de investigación el más empleado con el 85% de ellas es DSpace en ambos tipos de Repositorios, según sus estadísticas se ha instalado a nivel mundial en cerca de

2500 instituciones, información de acuerdo a su sitio web oficial⁶⁷. Su estructura es bajo un orden jerárquico que a continuación se muestra:

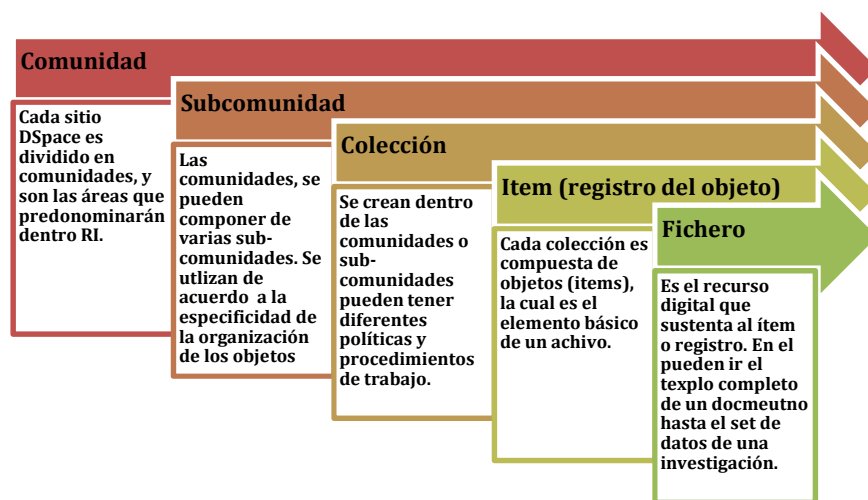


Figura 5. Estructura del repositorio con base en el software libre de DSpace (Elaboración propia)

2.7. Poblamiento de datos

¿A qué nos referimos con esta actividad? A la diseminación y revisión de los datos que se registrarán en el RI. Debemos tomar en cuenta las características propias del objeto digital en cuestión, así como su descripción detallada. Y analizar la inclusión de objetos complejos, de aprendizaje, de interacción, de datos o de análisis porque se combinarán nuevas entidades de información que deben ser claras tanto en la descripción así como en el acceso del objeto.

Los pasos que debemos tomar en cuenta en esta actividad es seleccionar el conjunto de datos que se van a poner a disposición, posteriormente corroborar que se puedan convertir a un formato abierto con el fin de facilitar el acceso, se debe aplicar una licencia abierta así como la verificación de la empaquetación, es decir una vez que se dispone del archivo de datos en un formato abierto, se debe embalar como toda la información necesaria para su acceso y consulta mediante herramientas de compresión de archivos e incluir el registro de su URI de acceso de modo que el recurso pueda ser identificado por su nombre fácilmente.

Este proceso se debe realizar a través de una integración de flujos de trabajo (workflow) por medio de dos procesos:

- La catalogación y marcaje de la información del objeto o del dato, así como el análisis de estos a través del formulario de la plataforma en la que se trabaje.

⁶⁷ LIRASIS (2019). *DuraSpace Registry*. Disponible en https://duraspace.org/registry/?filter_10=DSpace (Consultado 18-12-2019)

- b) La importación de los datos. La organización y el análisis se hace mediante un estructura de información, un marcaje vertical u horizontal según la naturaleza de los datos o también se puede optar por columnas y filas se tiene que considerar la naturaleza de los datos para que sea transparente el poblamiento y el registro quede 100% descrito en el RI.

La información que se manejará pueden ser de datos de y para la investigación, colaboraciones, comunicaciones científicas, sobre todo brindar el registro de los datos así como la descripción detallada de las características de su reutilización facilitando la redistribución e integración a otros conjuntos de datos. Una vez que se agreguen registros al RI, será muy complejo alterar o corregirlos por ser de estructura jerárquica.

Es necesario que ésta actividad esté a cargo de personal académico de la biblioteca, con el objetivo de que el contenido vaya avalado por una comunidad especializada en la catalogación y registro de la información bajo normas y lineamientos estandarizados ofreciendo registros de calidad. Es primordial elegir personal calificado que deposite los datos, la reputación de quien evalúe, revise y libere el contenido, será un indicador de la calidad de los contenidos e imagen institucional.

El poblamiento deberá ser acumulativo y perpetuo por lo que se debe establecer condiciones que mantengan activo el Repositorio mediante una política clara respecto a los contenidos, recordemos que el RI es un ente que tiene que ser alimentado para que ofrezca información puntual y actualizada.

2.7.1. Migración

Trabajar con los datos que se integrarán en el repositorio, tienen como principal finalidad ponerlos a disposición de toda la población de manera accesible, en formatos técnicos y legales que pueden ser publicados con ciertas características, para permitir que sean utilizados, reutilizados y redistribuidos.

Dentro del contexto de los Repositorios la migración puede realizarse mediante una ingesta de los datos para mantener su integridad por lo que se debe considerar los siguientes factores:

- Establecer el tiempo en el que se llevará la migración.
- La cantidad de duración de la inactividad requerida.
- Revisar la compatibilidad técnica.

Este tipo de migración implica mover los datos entre dos motores de base de datos. Sin embargo, cuando los datos originales cambian, afecta simultáneamente el protocolo y este se reflejará en la nueva interface.

Prácticas que se deben usar para proteger los datos durante una migración:

- Comprender qué datos se están migrando, dónde están, qué forma tienen y qué forma tomarán en su nuevo destino.
- Extraer, transformar y no duplicar datos (antes de moverlos).
- Implementar políticas de migración para que los datos se trasladen de forma ordenada.
- Probar y validar los datos migrados para asegurarse de que son precisos.
- Auditar y documentar el proceso de migración.

Sí los datos se encuentran en antiguas interfaces, plataformas que ya requieren actualización y se busca ponerlos a disposición de la comunidad, publicarlos y ofrecer visibilidad, será necesario llevar a cabo una migración como parte del proceso de implementación. Esta tarea involucra una doble acción primero se debe verificar la composición actual y segundo constituir y determinar la nueva estructura en la Interface elegida como destino final, por lo que se requiere una atención prioritaria en el momento de la transferencia de datos

Existen herramientas para el procesamiento y conversión de datos de fácil uso como el *Mr. Data Converter*⁶⁸ es un sitio que permite convertir un archivo de hoja de cálculo o de texto CSV a formato XML, JSON, etc. La forma de conversión es copiar los datos en el cuadro superior y seleccionar el formato; el resultado de la conversión se observa en el cuadro inferior del software. Además un punto a favor este tipo de sistemas son de acceso libre en la Web.

2.8. La estructura de despliegue

Los bibliotecarios cuando diseñamos un servicio, pensamos en el usuario que hace uso del mismo y cómo le agradecería consultar la información en pantalla, pero a veces esa visualización se encuadra en los elementos básicos de una ficha descriptiva de una obra o de un dato, por lo tanto la planeación en la estructura de despliegue de la información dentro del Repositorio debe estar ligada al tipo de usuario al que está dirigido, pero también considerar que los RI estarán abiertos a una amplia comunidad con diferentes niveles de escolaridad obligando al desarrollador del proyecto a preocuparse en una estructura que ofrezca diversas alternativas en el despliegue y consulta de los recursos ingresados.

En este caso, la estructura del despliegue de los datos debe tener una visión amplia enfocada a distintos tipos de usuarios y sobre todo a que los datos sean visibles y útiles para cada usuario no importando el nivel de búsqueda y escolaridad. Por ello es uno de los puntos clave para dar a difusión al trabajo que se realiza en la implementación del Repositorio.

Para ofrecer una composición clara y fundamentada para el despliegue de la información almacenada se debe trabajar en tres líneas, las cuales han sido

⁶⁸ Data Converter. Convierte datos en varios formatos compatibles con la web, incluidos HTML, JSON y XML. <https://shancarter.github.io/mr-data-converter/>. (Consultado: 28-06-2019).

tomadas con base en la experiencia personal que se ha tenido en el desarrollo de los RI del área de Ciencias de la Tierra en la UNAM, siendo estas:

1. Establecer un despliegue de los principales elementos y metadatos que permitan identificar el registro y el recurso ligado con base en la iniciativa empleada, se recomiendan seis principales datos a desplegar son: autoría, títulos del recurso, fechas de creación, ingreso y almacenamiento, temas y/o palabras clave y liga al recurso. Después de estos elementos, los que se incluyan posteriormente son importantes pero dependerá de cada institución que otros elementos quieren mostrar en una primera pantalla de despliegue y aumentar su número en una segunda pantalla.
2. La estructura del contenido del Repositorio se basa en comunidades y subcomunidades o carpetas y subcarpetas, por lo que es necesario considerar la visualización del despliegue y la clasificación de los datos. En el caso del Repositorio de literatura sus divisiones están estructuradas por lo regular por el tipo de documentos (libros, capítulos, artículos, tesis, reportes, etc.) y en subcomunidades mediante una división por año o temática. A diferencia del Repositorio de datos su estructura es compleja por el tipo de metadatos empelados y en la estructura de comunidades por lo regular son de acuerdo a la especialización de los datos incluidos (taxonomías, familias, tipologías, por mencionar algunos). En este sentido la estructura de despliegue debe ser considerada con base en el área de especialización.
3. Se debe entender que un Repositorio no es una base de datos sino la puerta para ofrecer la ciencia abierta al mundo académico, por ello al estructurar el despliegue de los datos, debe estar basado en lo que necesitan los diversos usuarios que accederán, pero también es preioritario dar a difusión a los Autores para ofrecer una visibilidad, citación e impacto entre los usuarios y con ello aumentar la presencia de la institución en el medio académico a nivel mundial. Es importante establecer una media entre las metaetiquetas básicas de despliegue con las que conllevan los elementos de visibilidad institucional y que son necesarios para contar con citación académica.

Con base en lo anterior, la estructura del despliegue es trabajo académico que debe ser desarrollado por bibliotecarios, los Autores y/o responsables de la información y las autoridades institucionales para contar con diferentes visiones y al mismo tiempo establecer claramente lo que requeriría el usuario en la primera consulta de la información en el Repositorio.

2.9. DSpace: su estructura

Se considera este software libre por el alto uso y demanda que tienen actualmente dentro del desarrollo de los Repositorios y con ello facilitar la conservación y la

difusión en Acceso Abierto al contenido digital como texto, imágenes estáticas y en movimiento, videos, MPEG, conjuntos de datos así como otros formatos permitiendo una interoperabilidad.

Es una plataforma de código abierto licenciado bajo *BSD open source license*, significa que cualquier organización puede usar, modificar e incluso integrar el código en su aplicación comercial sin pagar ninguna tarifa de licencia.

La estructura que a continuación se presenta es con base en un contexto bibliotecológico con el fin de conocer la parte técnica en donde el personal académico que labora en la biblioteca y/o el que desarrollará un Repositorio debe tomar en cuenta con el fin de establecer permisos, licencias y configurar el funcionamiento de esta plataforma.

La arquitectura de DSpace se divide en 3 principales grupos:

- **Módulos de Aplicación:**
 - XMLUI: frontend basado en Manakin y Apache Cocoon.
 - JSPUI: frontend en servlets y JSPUI.
 - REST: API para CRUD y otras operaciones sobre HTTP.
 - OAI: data-provider OAI-PMH para diseminación del contenido del repositorio.
 - SWORD: servidor SWORD para ingesta de contenidos al Repositorio desde otros sistemas.
 - SOLR: motor de indexación → cores: oai, search, statistics
- **De lógica de negocio**, basada en servicios como MetadataValueService y DSpaceObjectService, entre los más usados.
- **De Almacenamiento Interfaces Capa de Acceso a Datos (DAOs)**, objetos de base de datos, base de datos específica y almacén de *bitstreams*.

Las API se dividen en 3 capas: de Servicios, de Acceso a datos y de Objetos persistentes.

En cuanto a los módulos de contenido y de la administración del software que apoyará en la organización del mismo:

- **Módulos de contenido:** se organizan en una o más comunidades de nivel base y los elementos son:
 - Comunidades organizadas jerárquicamente.
 - Subcomunidades y colecciones, estas últimas son los “estantes”, que agrupan contenido.
 - Ítems: son las obras que van en los “estantes” y que se pretende que el público encuentre.
 - Los metadatos refieren las obras mediante datos estructurados que describen otros datos, lo que llámanos *datos sobre datos*, incluye varios tipos.

- Gestión de autoridades, ya sea abiertas o cerradas, tesauros, sistemas de clasificación, taxonomías, lista de encabezamientos de materia, bases de datos de investigadores, jerarquía de instituciones, etc.
 - Identificadores persistentes permiten identificar un recurso de manera independiente a la dirección del repositorio o sitio donde dicho recurso se encuentra.
 - Los *bitstreams* son representaciones digitales, totales o parciales, de los ítems.
- **Módulo de autorización – Permisos**
 - Persona o grupo autorizado
 - Acción permitida
 - Sobre qué objeto
 - Desde cuándo y hasta cuándo.
 - Grupo Quién representado por la entidad E-Person, corresponden con los usuarios registrados en el sitio.
 - Grupos predefinidos, existen 2 grupos de usuarios por defecto:
 - ✓ Anonymous: representa a todos los usuarios, autenticados o no, y/o, en su caso, cualquier usuario con pertenencia a este grupo, independientemente de su pertenencia a otros grupos.
 - ✓ Administrator: usuarios que son administradores del sistema y representa el rol de SUPERUSUARIO, con autorización plena sobre DSpace.
 - Grupos de trabajo son usados para configurar los flujos de trabajo entre los que se encuentran:
 - ✓ ADMIN: administrador de elementos.
 - ✓ ADD: agregar publicaciones.
 - ✓ REMOVE: eliminar publicaciones.
 - ✓ WRITE: Modificar datos.

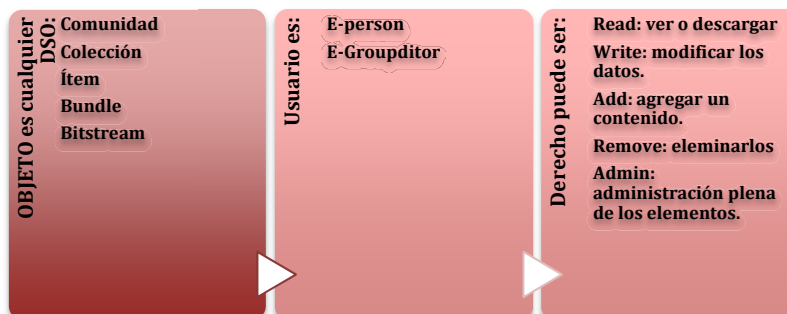


Figura 6. Por defecto, todas las comunidades, colecciones e ítems tienen permiso de READ para el grupo Anonymous. (Elaboración propia)

- **Estadísticas.** En este módulo se registran datos de accesos, descargas, navegación de los contenidos en un índice Solr. A diferencia de herramientas como Google Analytics y PIWIK, los registros se crean por cada acceso al servidor y no desde un agente de usuario. También provee estadísticas en vivo en acceso a ítems por origen y fecha, en las descargas de *bitstreams* por origen y fecha; finalmente, en los ítems más accedidos del Repositorio.
- **Servicios para usuarios.** Este se basa en la exploración desde de los metadatos: autores, temas, años, etc., y a partir de su organización lógica de comunidades, subcomunidades y colecciones que se establecieron al inicio del proyecto y, finalmente, es la presentación web.

2.9.1. El diseño final

En todo proceso de desarrollo de un Repositorio debe existir un diseño final, no de imagen, sino de contenido, con base en la distribución de las metaetiquetas y el acomodo de la información en comunidades y subcomunidades, hasta llegar al recurso de despliegue. Por ello, este elemento es fundamental debido a que una vez definido y diseñada su composición no podrá ser modificada sino reestructurada mediante un proceso de cambios constantes en el propio Repositorio.

El diseño final debe basarse en los elementos señalados de los puntos anteriores, los cuales quedarían resumidos de la siguiente manera:

- a. Información o datos a incluir, lo que sin duda es el punto clave del diseño final, dependerá del software a utilizar. Debemos entender que de ello dependerá el futuro del Repositorio. Los datos para la investigación son ejemplo especial para establecer el diseño final en comunidades y subcomunidades, ya que los recursos que se visualicen se mostrarán a la comunidad y marcarán las líneas del software y los metadatos a emplear.
- b. En el caso del Repositorio de Datos, las opciones de software son variadas debemos perseguir la línea que marcan las instituciones que rigen la ciencia abierta en el país para que ser interoperables. Por ello, el diseño final debe contemplar parámetros internacionales de intercambio de información con software libre, pero sobre todo, reconocidos por organismos normativos de aceptación mundial que avalen su desarrollo y continuidad.
- c. Un elemento más en el diseño final son los metadatos empleados, ya que, al utilizar aquellos que conforman una estructura básica para el tipo de contenido que se maneje en el Repositorio, podemos caer en el problema de que no se desplegará lo determinados mostrar y posiblemente hasta no incluir toda información que deseemos, esto es la limitante de algunas metaetiquetas cuando no son incluidas en el tipo de

metadato que utilizamos. El diseño no solo debe contar con la posibilidad de haber establecido claramente una estructura bien documentada del proyecto, sino que además es necesario hacer cambios constantes en su crecimiento; por ello, el establecimiento de las comunidades y sus subcomunidades con las metaetiquetas bien definidas ayudarán a que la estabilidad del proyecto se consolide.

El diseño final de una propuesta de Repositorio debe contar con la participación directa de los autores y responsables de la información, las autoridades institucionales, los bibliotecarios académicos y el personal de cómputo que en su conjunto deben trabajar en conjunto para lograr un resultado de consulta real de la información hospedada en el Repositorio. El trabajo, tiempo y recursos incluidos en su desarrollo son importantes y en algunos casos sobrepasa las expectativas de sus creadores. Por otra parte, la interacción con los miembros que promueven la ciencia abierta en el país no puede quedarse de lado (se reciban o no recursos económicos de su parte) para que, en el caso de México, podamos ser cosechados por el Repositorio Nacional.

2.9.2. Diseño web

Una vez determinado el diseño final de la estructura interna del Repositorio, es necesario considerar la imagen institucional que se asumirá y permita establecer una marca en el gremio académico; por ello, el diseño web es primordial para lograr una identificación inmediata en el momento de su consulta.

El humano, por naturaleza, busca mostrar varios elementos que se relacionen con las actividades que realiza; por ello, se agregan fotos, logos y además de una serie de colores que nos identifiquen como persona o institución para obtener un posicionamiento entre la comunidad que acude a nosotros; sin embargo, para un Repositorio que estará disponible para ser consultado a nivel mundial se debe valorar el diseño web que muestre su contenido, de dónde proviene y quién lo desarrolló, sin perder su esencia de difundir la ciencia abierta.

En este sentido, la participación –aparte de las personas señaladas en el punto anterior– de un diseñador web y gráfico son elementos preponderantes para que el resultado sea de lo más satisfactorios. En el caso de no contar con este recurso, los miembros que tomen decisiones tienen la tarea de no saturar la imagen web, con la obligación de ofrecer la imagen institucional y de contenido agradable a la vista del usuario, además de intuitiva.

Mostrar una imagen institucional y del contenido, forma parte del éxito de Repositorio. Asimismo, es importante usar imágenes gráficas en el diseño, pero es muy recomendable no poner movimiento los logos ni cambios de color o anuncios, pues es una herramienta para la difusión de la ciencia abierta y no un sitio web de divulgación o de información institucional.

El diseño web es en sí la imagen que percibirá el usuario final al momento de hacer la búsqueda. Por ello, no debemos generar distractores en su línea de búsqueda en el Repositorio, se debe concentrar en lo que requiere y en la información que necesita recuperar, facilitando con ello la recuperación de datos o recursos buscados y permitiendo, con ello, una mayor visibilidad, impacto y presencia de nuestros contenidos entre la comunidad científica y académica que lo consulte.

CAPÍTULO TRES ELABORACIÓN DE UN REPOSITORIO DE DATOS PARA LA INVESTIGACIÓN EN CIENCIAS DE LA TIERRA

En este capítulo se presenta el desarrollo de un Repositorio de Datos con base en la normatividad y lineamientos del Conacyt y otras instituciones de carácter internacional, por el tipo de datos que contiene y con la finalidad de sumarse a un Proyecto Nacional, contribuyendo, impulsando y difundiendo la ciencia abierta tanto en México como a nivel internacional.

Deseamos se posicione como uno de los primeros Repositorios de consulta y calidad de contenidos, empleando una serie de elementos que preserven los datos para posteriores investigaciones y genere nuevas líneas de acción y de conocimiento. Con esto, intentamos fomentar la ciencia e investigación con la consulta de recursos de información diseminados y especializados, mediante herramientas que posibilitan canales para compartir datos a través de plataformas y así poderlos reutilizar.

Dentro del contexto de las Ciencias de La Tierra, los datos se generan a gran escala y con ellos se informa a la sociedad y a los investigadores sobre el impacto que tiene la naturaleza sobre La Tierra, la Ciencia Espacial y el Mar, por lo que investiga soluciones para la sostenibilidad del planeta, así como la mitigación del impacto de los desastres naturales que afectan a las poblaciones con el fin de poder abordar y ofrecer soluciones a los desafíos sociales que van desde evaluar y responder a los peligros naturales, el cambio climático, el uso de la energía, los recursos naturales, así como estudiar y analizar el comportamiento de los océanos, el aire y la tierra. Por ello, un proyecto de este tipo busca ser una herramienta que pueda albergar gran cantidad de datos que se producen en los diferentes Servicios Nacionales de México, como el Sismológico, Mareográfico, el Magnético, Solarimétrico, el de Geodesia Satelital y Clima Espacial, por ejemplo.

Pero todo ello debe hacerse bajo una normatividad, directrices y políticas, así como con estándares internacionales que cubran el valor de los datos abiertos, y compartirlos bajo ciencia abierta. Debemos intentar que el uso sobre datos de Ciencias de La Tierra, el Mar y la Atmósfera garantice y mantenga una sociedad orientada e informada, con líneas de protección civil, conservación y estabilidad de datos.

Llevar a cabo el diseño de herramientas de alcance internacional facilita la preservación de los datos, ya sea en plataformas de enlace, interfaces, software, medios locales, en línea, de acceso a los datos, así como en la nube. Con ello, podemos garantizar la disponibilidad de los datos y al mismo tiempo una preservación de los contenidos que ofrezcan una perdurabilidad.

Así, a continuación presentamos una serie de elementos que, con base en la experiencia con que se cuenta y la literatura consultada, debe ser considerada por aquellos que están estructurando un Repositorio Institucional (RI) para la difusión y publicación de su ciencia de manera abierta, desde un contexto bibliotecológico. Se diseña en particular para el área de las Ciencias de La Tierra con la finalidad de promover el intercambio de datos y cooperar con otras instituciones en el monitoreo e investigación a nivel nacional e internacional de las diversas

actividades sobre La Tierra, el Mar y la Atmósfera. En él se puede albergar información especializada en diversas formas y formatos, que van desde los datos puros tanto de como para investigación, además de publicaciones, informes, videos, mapas, fotos, etcétera, logrando ponerlo a disposición de la sociedad y apoyar el desarrollo científico del país.

Es importante mencionar que la metodología que se describe a continuación sobre la propuesta de creación de un Repositorio de Datos es con base en la normatividad y lineamientos del Conacyt y de algunos estándares internacionales asociados por la especialidad de la información, buscando interoperar con el Repositorio Nacional y sumarnos a un Proyecto de alcance Nacional de Repositorios en México para proyectar la ciencia abierta.

Por otra parte, se ofrece una revisión en cuanto a la estructura técnica e información registrada en los 96 Repositorios Institucionales tanto de Datos como de Literatura, a fin de presentar un panorama más amplio y mostrar los elementos que se pueden considerar en estos dos tipos de fuentes con el propósito de generar una herramienta que ofrezca información detallada y apoye su diseño basado en las necesidades de los diferentes tipos de usuarios.

3.1. Planeación

Bajo este entorno, se planea el diseño y puesta en marcha de un Repositorio de Datos en el área de Ciencias de La Tierra, en donde puedan manejarse diversos datos de recursos geotérmicos, movimientos sísmicos, de clima espacial, entre otros, los cuales incidirán directamente en el impacto y producción de Ciencias de La Tierra manteniendo a una sociedad informada y generando datos para nuevas investigaciones de los eventos naturales a los que estamos propensos en México, mediante proyectos encaminados a la difusión ciencia abierta.

Para poder llevar a cabo el RI de datos, se debe hacer un análisis completo y detallado del tipo de información que se incluirá y publicará, por lo que es necesario estructurar un grupo de trabajo que realice dicha actividad con resultados importantes, quedando la propuesta conformada de la siguiente manera:

- a) Especialistas de los datos, investigadores o académicos responsables de los datos con los que se trabajarán, ya sean generados de forma directa o por medio de la administración de tecnologías relacionadas.
- b) Personal académico bibliotecario, quienes manejan la estructura técnica de los metadatos tanto Dublin Core como OpenAIRE y DataCite, para estructurar de forma eficiente la descripción física de cada dato y sus formas de recuperación.
- c) Personal de cómputo que se encargará de la instalación de servidores, software y configuración de las plataformas, logrando con ello ejecutar una estrategia de trabajo que permita un crecimiento constante del proyecto.

Para ello se trabajó en una tipificación de los datos, que se manejan en el área de Ciencias de La Tierra, con la finalidad de entenderlos y estructurarlos dentro del Repositorio:

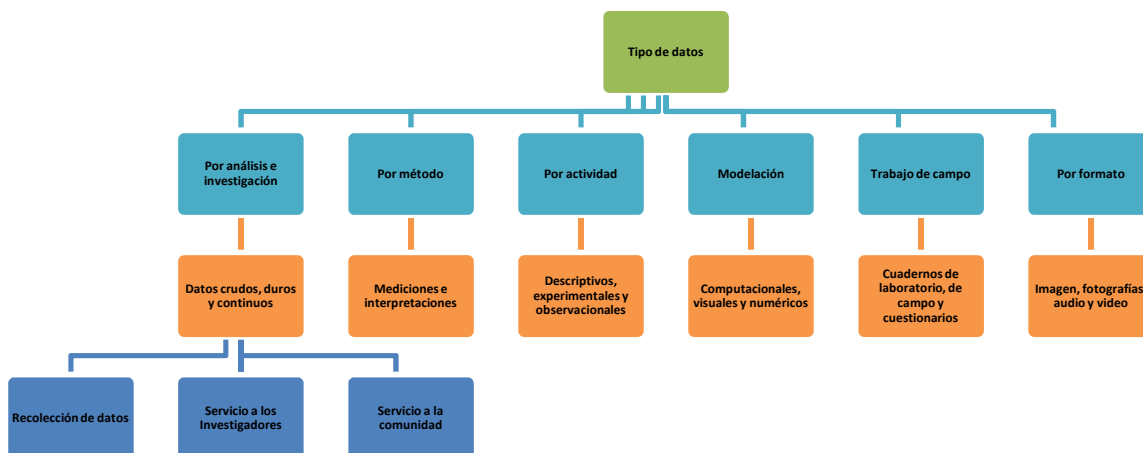


Figura 7. Tipo de datos que se trabajan en el RI.

El propósito del Repositorio de Datos es la compilación, organización, conservación y difusión de los datos científicos pero además sí se considera necesario incluir producción académica ésta debe registrarse y cumplir con las normas y estándares correspondientes a su contenido. Y poder establecer una contestación con base en su prefijo de metadatos –debiendo ser *Datacite para los datos* y *Dublin Core* para la literatura, según sea el caso– además de la programación técnica adecuada para llevar a cabo la indización, recuperación e interoperabilidad para su libre acceso y despliegue de contenidos.

Con el propósito de almacenar datos científicos, ya sea de servicios, proyectos en colaboración a nivel nacional o internacional que se pueda compartir y/o rehusar de manera libre y estructurada bajo las condiciones de Acceso Abierto y/o ciencia abierta además de cumplir las disposiciones establecidas en la legislación y normatividad internacional.

También es necesario establecer lineamientos para la conservación de los datos en el Repositorio para el futuro de la investigación sobre los temas de la especialidad, logrando así cubrir tres importantes propósitos: su organización, difusión y conservación a mediano y largo plazo.

En ese sentido, la planeación la determinamos estableciendo siete etapas que van desde el análisis y la organización de la datos, pasando por su revisión y permisos del uso de los datos, así como sus diversos licenciamientos, además de la definición del identificador de los autores, la estructura de los metadatos que se

deben de aplicar en el RI para lograr una cosecha⁶⁹ (interoperabilidad) exitosa, tomando en cuenta además el diseño web, el despliegue de los datos, la imagen institucional y la liberación para su interoperabilidad y consulta.

A continuación presentamos un esquema de cada una de las etapas por las que se atraviesa en el desarrollo de un Repositorio de Datos, desde su creación, interoperabilidad con el agregador del Repositorio Nacional, hasta colocarlo a disposición y consulta de los usuarios:

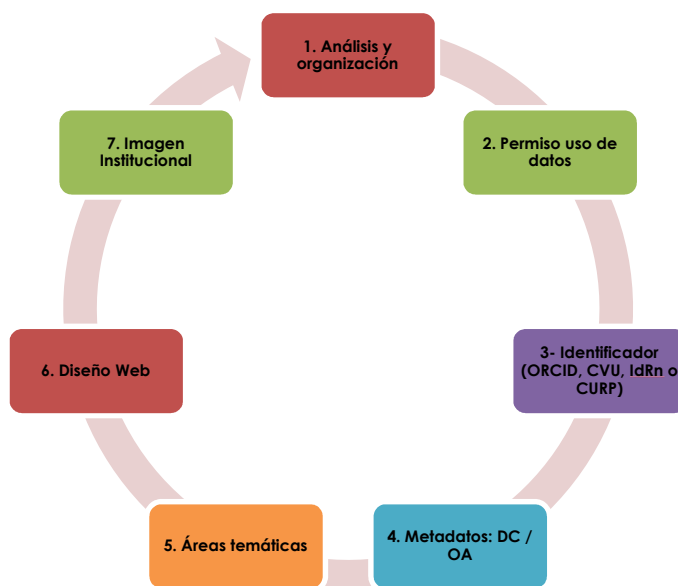


Figura 8. Etapas de la planeación para el desarrollo de un Repositorio de Datos.

Para llevar a cabo estas etapas de la planeación, es necesario dividir las en actividades técnicas, académicas y administrativas, mismas que se designan a cada personal implicado en el proyecto y se describen a continuación:

Técnicas:

- Instalación de los sistemas operativos, de DSpace y software que se requiere para desplegar la información en el RI.
- Configuración de los servidores y del equipo de cómputo que se utilizará para hospedar el RI.
- Programación y diseño del formulario para la asignación de metadatos bajo la normatividad de OpenAIRE y DataCite para cada uno de los formatos documentales a incluir en el Repositorio.
- Configuración del uso del prefijo de los metadatos para Datos bajo Datacite.
- Enlace con los Catálogos de Autoridad con el Repositorio Nacional, llamados servicios REST.

⁶⁹ Proceso que confirma que las estructuras técnicas están normalizados bajo un estándar y son interoperables entre los Repositorios. Su función principal es mover metadatos de un servidor local a un agregador con la finalidad de reunirlos con otros metadatos y proporcionar un servicio de valor agregado. Es importante mencionar que se requiere de un acuerdo entre el proveedor de metadatos y el cosechador.

- Instalación y software especializado para el despliegue de información, este varía de acuerdo con área de Ciencias de La Tierra que se trabaje.
- Configuración de los protocolos y verificación de la interoperabilidad de RI con el Repositorio Nacional, así como con otras Instituciones.

Académicas:

- Análisis de los datos y en caso de incluir objetos digitales, se deberán considerar los elementos adicionales para la descripción e interoperabilidad.
- Descripción según la naturaleza de los datos. Pueden ser sismológicos, geomagnéticos, geotérmicos, de clima espacial, entre otros.
- Parametrización de DSpace y del registro Handle (identificación y personalización del URL) para la colección con base en las diferentes comunidades y colecciones que se trabajarán para la administración del sistema.
- Revisión y análisis de la documentación de la responsabilidad de derechos de uso y manejo de los datos y tipo de licencias.
- Definición de los identificadores de los autores por medio de ORCID, CURP, CVU, DNI o RFC, según se requiera.
- Implementación de las áreas temáticas del Conacyt, así como las necesarias para la clasificación de la información.
- Diseño del sitio web que permita la visibilidad del proyecto en general y del Repositorio en particular, para su difusión y consulta a nivel nacional e internacional.
- Inclusión en los descubridores de uso público para su fácil recuperación en los primeros registros de búsqueda en la Web, por medio del etiquetado META para su indización en los buscadores en Internet.

Administrativas:

- Proceso de cotizaciones del equipo de cómputo para realizar el proyecto.
- Administración de las partidas presupuestales para ejercer las compras requeridas.
- Colaboración en la realización de los informes de desarrollo del avance del proyecto.
- Coordinación para realizar las salidas y visitas a las instituciones que colaborarán en la aportación de datos y/o material documental para la inclusión en el RI.
- Realizar los pagos al personal de servicio social y de ayudas a los estudiantes que apoyen en el RI.

A continuación se presenta el modelo de planeación general del desarrollo de tres Repositorios Institucionales de Datos que se aplicó en la Biblioteca Conjunta de Ciencias de La Tierra (BCCT). Este modelo ofrece los elementos necesarios a considerar que se deben cumplir a partir de llevar a cabo a partir del registro de la *Convocatoria para desarrollar Repositorios de Ciencia Abierta* –que se presenta

en Conacyt–, el diseño y las etapas que pueden apoyar y orientar en el desarrollo de este proyecto.

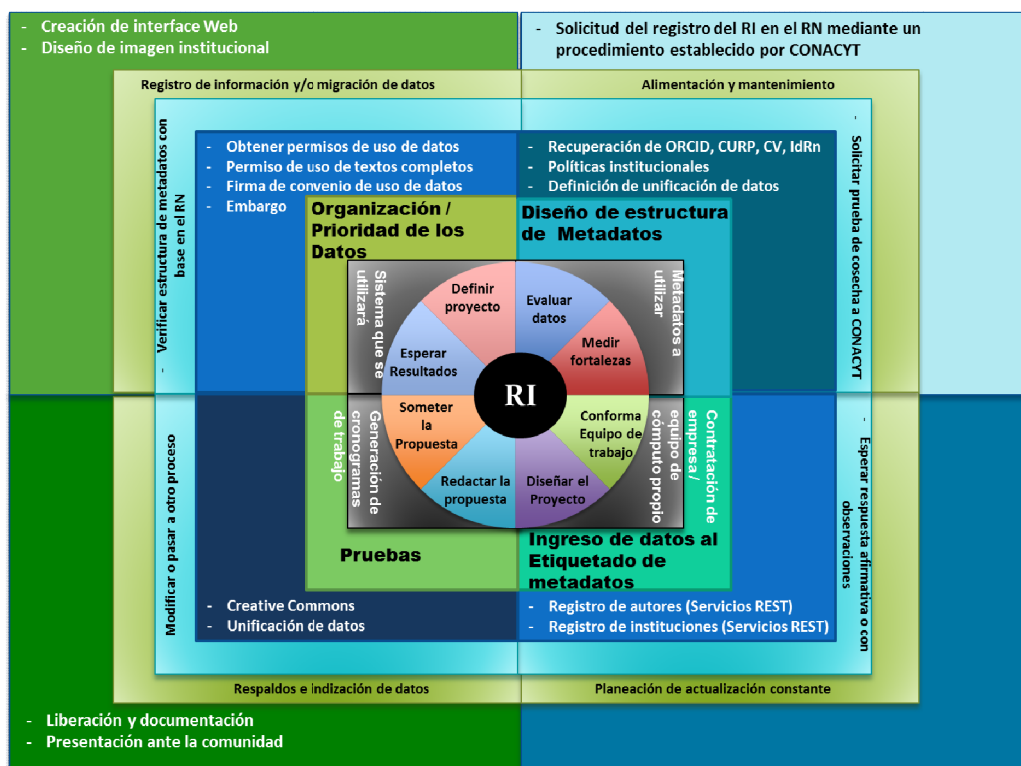


Figura 9. Modelo de planeación general en el desarrollo de un Repositorio Institucional de Datos que ofrece los elementos a considerar a partir de la *Convocatoria para Desarrollar Repositorios de Ciencia Abierta*. El modelo fue elaborado por el personal académico de la BCCT.

3.2. Contenido

El tipo de contenido que se incluirá en el Repositorio de datos debe funcionar bajo el objetivo institucional, así como la misión por la cual se crea. Debemos partir del análisis, revisión y gestión, con el apoyo de los especialistas en los datos. Es importante considerar que esta herramienta busca apoyar a la investigación y la docencia con datos que pueden abarcar desde contenidos duros –como números, caracteres, imágenes, representaciones físicas en símbolos, generalmente presentados en forma de tabla, es decir cada columna representa una variable particular y cada fila una asignación de valores para cada variable de uso–, así como su posible recolección, análisis e inclusive sus interpretaciones para un posterior tratamiento. En este punto es importante incluir metadatos de la especialidad. Así que se debe crear un Repositorio Activo con poblamiento continuo, basado en datos bajo principios FAIR que cumplan con los estándares de búsqueda, accesibilidad, interoperabilidad y reutilización.

La base primordial es incluir los datos puros, crudos y otros obtenidos como son las salidas de campo, el análisis y las relaciones, el uso y la interpretación de los

datos, y en caso de ser necesario publicaciones, informes técnicos, cartas geográficas, videos, mapas, entre otros que estén relacionadas con los datos.

Se debe trabajar de manera conjunta con los especialistas (investigadores, técnicos académicos, docentes, estudiantes de posgrados, etc.) para conocer la estructura de los datos según el área de Ciencias de La Tierra. Algunos ejemplos de ello son:

- *Geotermia*: balance iónico, química del agua, temperatura, manifestaciones, coordenadas geográficas, metodología, muestreo, estado, municipio, localidad, entre otros.
- *Sismología*: datos de aceleración, velocidad, desplazamiento, profundidad, altitud, magnitud, epicentro, de tiempos, de redes y estaciones sismológicas.
- *Clima espacial*: índice geomagnético, fechas de publicación, flujo de rayos, índices Dst y K Local, mediciones, datos en tiempo real, temperatura, frecuencia y latencia.

Es importante mencionar que se debe trabajar con datos producidos por las entidades académicas que posean la responsabilidad de los derechos de autor, un ejemplo de ello es que contamos con la autoría de los datos de algunos Servicios Nacionales (estaciones y redes), por lo que se cuenta con derechos de autor, el uso y manejo así como la producción continúa e ininterrumpida de los datos, logrando con ello conformar un Repositorio de datos de alto impacto ante los investigadores y la sociedad en general.

Dentro del campo jurídico de los derechos de autor y licenciamiento, nos basamos en los acuerdos y principios de la ciencia abierta y en las políticas y lineamientos del Conacyt, por ser el que encabeza el *Proyecto de Repositorios del Consejo Nacional de Ciencia y Tecnología sobre Acceso Abierto de la información científica y tecnológica*. En cuanto a derechos de autor, consideramos los siguientes puntos, siendo prudente aquí aclarar que los alcances de los derechos de autor son vastos, pero para efectos de este trabajo solo consideramos los que se analizaron bajo alcances de Conacyt y la ciencia abierta:

- Analizar y determinar la creación de los datos. ¿De quién son los datos? ¿A quién le pertenecen los datos?.
- Establecer los límites para poder ofrecer las autorizaciones correspondientes según lo determine el creador o propietario.
- Buscar una protección adecuada según la naturaleza de los datos.
- En caso de incluir datos de otras instituciones ajenas, considerar las cartas y autorizaciones escritas, ya sea para sesión de derechos y/o determinar cómo se establecerá su derecho de autor en la inclusión del Repositorio.
- Lo que se incluya en el RI deberá analizarse y verificar que cumpla con los lineamientos establecidos para no incurrir en algún problema legal.

En cuanto al licenciamiento se optó el uso de la *Licencia Creative Commons (CC)*, con el fin de establecer bajo que condiciones de uso pueden utilizar los datos, respetando los derechos y obligaciones de la licencia que se escoja.

Se revisan las versiones y actualizaciones de las CC. Considerando el alcance con los nuevos formatos, se decidió trabajar con la versión 4.0 bajo los términos de *Licencia Atribución-NoComercial-SinDerivar 4.0 Internacional (CC BY-NC-ND 4.0)*, es decir, para el uso de los datos que se producen y puedan ser consultados por cualquier usuario, quedando de la siguiente manera:



Figura 10. Licencia Atribución-NoComercial-SinDerivar 4.0 Internacional (CC BY-NC-ND 4.0)⁷⁰

En esta licencia puede compartir, copiar y redistribuir el material en cualquier medio o formato. El licenciador no puede revocar estas libertades mientras cumpla con los términos de la licencia, bajo las condiciones siguientes:

- *Reconocimiento*: debe reconocerse adecuadamente la autoría, proporcionar un enlace a la licencia e indicar si se han realizado cambios. Puede hacerlo de cualquier manera, pero no que sugiera que tiene el apoyo del licenciador o lo recibe por el uso que hace.
- *NoComercial*: no puede utilizar el material para una finalidad comercial o con fines de lucro.
- *SinObraDerivada*: en caso de remezclar, transformar o crear, no puede difundir el material modificado.

Como se menciona en el capítulo 2 hemos definido dos tipos de Repositorios de datos, el de Datos para la Investigación que contiene datos puros y continuos de apoyo a investigaciones o para investigar nuevas propuestas y tienen la condicionante de que sus datos pueden ser reutilizados en diferentes disciplinas; y el de *Datos de Investigación*, perteneciente a una infraestructura académica en donde los recursos y servicios utilizados por la comunidad científica realiza investigaciones de datos incluidos y vinculados en una publicación (artículo, libro, capítulo, tesis, etc.). Para este último, que requieren los datos su publicación en una Revista académica, se toman en cuenta los permisos editoriales de *Sherpa Romeo*, con el fin de mostrar los derechos de autor y las políticas de autoarchivo de Acceso Abierto así como todas sus condicionantes.

Es importante revisar las políticas editoriales que tiene cada revista, para la publicación de los datos.

⁷⁰ La versión 4.0 del paquete de licencias CC se publicó el 25 de noviembre de 2013, luego de que CC realizará un proceso de versiones de varios años. Las primeras discusiones públicas de 4.0 se llevaron a cabo en la Cumbre Global 2011 de CC.

Se presenta un ejemplo de una revista arbitrada e indizada, que en su portal comunica sobre sus políticas editoriales tanto de edición de documentos como las características de los datos para publicar y otros materiales subyacentes bajo acceso abierto y de reuso:

Figura 11. Condiciones de la publicación en la revista.⁷¹

Por lo tanto, todo el contenido de las Revistas *ANGEO* y *ANGEOD* se distribuye bajo la licencia internacional Creative Commons Attribution 4.0 (CC BY 4.0) y los autores conservan los derechos de autor.

Figura 12. Política de la publicación de los datos.⁷²

En las políticas de publicación hablando específicamente de los datos deben estar acompañados de su información complementaria además de una sustentación de los hallazgos de la investigación para que sean localizables, accesibles, interoperables y reutilizables (principios FAIR), no solo de lectura para los humanos, sino también para las máquinas.

⁷¹ *Annales Geophysicae. An interactive open-access Journal of the European Geosciences Union*. Disponible en https://www.annales-geophysicae.net/about/publication_policy.html. (Consultado: 8-07-2019).

⁷² *Annales Geophysicae. An interactive open-access Journal of the European Geosciences Union*. Disponible en https://www.annales-geophysicae.net/about/data_policy.html. (Consultado: 8-07-2019).

Annales Geophysicae
An interactive open-access Journal of the European Geosciences Union

EGU.eu | EGU Publications | EGU Highlight Articles | Contact | Imprint | Data protection |

Submit a manuscript | Manuscript tracking

Metadata and full-text XML harvesting

Interested parties aiming to access the metadata and full-text XML files of the articles published in Copernicus' journals are invited to use our OAI-PMH interface. The OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) is a low-barrier mechanism for repository interoperability used to harvest (or collect) the metadata descriptions and in our case also the full-text articles.

Copernicus Publications fully supports the OAI Metadata Harvesting Protocol 2.0. Metadata for all published articles is available in the Dublin Core format (oai-dc). In addition, we provide full-text XML using the NLM format (pmc). The latter is available for all articles published after November 2014.

Examples of OAI-PMH requests

- To retrieve information about the repository: <https://oai-pmh.copernicus.org/oai.php?verb=Identify>
- To retrieve the set structure: <https://oai-pmh.copernicus.org/oai.php?verb=ListSets>
- To retrieve the available metadata and full-text formats: <https://oai-pmh.copernicus.org/oai.php?verb=ListMetadataFormats>
- To retrieve records: https://oai-pmh.copernicus.org/oai.php?verb=ListRecords&metadataPrefix=oai_dc
- To retrieve an individual metadata record (e.g. aab28576): https://oai-pmh.copernicus.org/oai.php?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:publications.copernicus.org:aab28576 for oai-dc or <https://oai-pmh.copernicus.org/oai.php?verb=GetRecord&metadataPrefix=pmc&identifier=oai:publications.copernicus.org:aab28576> for pmc

Search articles
Search
Author

Search web pages
Search

Figura 13. Política de la publicación de los datos: Metadatos y XML.⁷³

Así pues, es importante poner a disposición el acceso a los metadatos y archivos XML del texto completo de los artículos publicados, con la finalidad de llevar a cabo la interoperabilidad con otros repositorios de interés. Los metadatos en esta caso de los artículos publicados están disponibles en el formato Dublin Core (oai-dc) y con el de la NLM (pmc).

3.3. Diseño del DSpace para el Repositorio

Como se comentó en el capítulo anterior, este software de código abierto es una herramienta para la administración de colecciones digitales y soporta una gran variedad de objetos, entre ellos los datos.

El diseño que se consideró se basó en las necesidades de la información que se persigue ofrecer en el RI de datos. También se tomaron en cuenta la tipología de los datos, el proceso de cómo se quieren hacer llegar y cómo se llevaría a cabo el despliegue de la información y su uso, pues son parte esencial del diseño, mismo que nos ayuda a definir el proyecto y su futuro.

A continuación presentamos las etapas del diseño, con base en nuestra experiencia, así como la descripción de cada una de ellas:

⁷³ *Annales Geophysicae. An interactive open-access Journal of the European Geosciences Union.* Disponible en https://www.annales-geophysicae.net/about/xml_harvesting_and_oai-pmh.html. (Consultado: 8-07-2019).

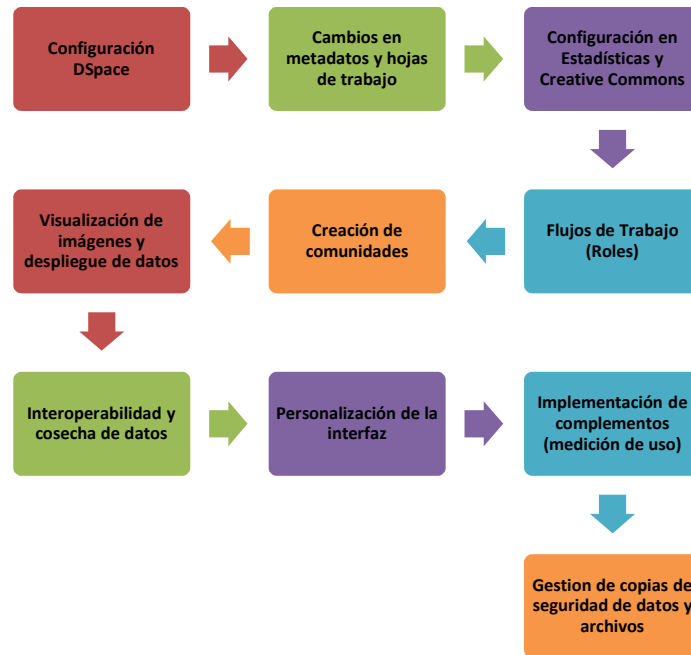


Figura 14. Etapas en el diseño de DSpace.

Para ello se desglosa, gráficamente, el desarrollo de las actividades que se llevarán a cabo en DSpace y que pueden tomarse en cuenta y lograr una implementación óptima de un Repositorio de Datos:

Diseño Técnico de Cómputo

- Instalación y configuración del DSpace, direcciones de acceso IP, el Handle del Repositorio y la Interfaz JSPUI (ésta última es por la que se decide salir en la presentación de los datos y el enlace a las diferentes estructuras y plataformas con las que conectarán los datos).

```

[INFO] DSpace XML-UI (nanakin) :: Local Customizations ... SUCCESS [ 8.182 s]
[INFO] DSpace DSP-UI ... SUCCESS [ 5.141 s]
[INFO] DSpace JSP-UI :: Local Customizations ... SUCCESS [ 5.141 s]
[INFO] DSpace RDF ... SUCCESS [ 3.050 s]
[INFO] DSpace RDF :: Local Customizations ... SUCCESS [ 4.291 s]
[INFO] DSpace REST :: API and Implementation ... SUCCESS [ 20.394 s]
[INFO] DSpace REST :: Local Customizations ... SUCCESS [ 4.111 s]
[INFO] DSpace SWORD ... SUCCESS [ 1.080 s]
[INFO] DSpace SWORD :: Local Customizations ... SUCCESS [ 3.601 s]
[INFO] DSpace SWORD v2 ... SUCCESS [ 28.497 s]
[INFO] DSpace SWORD v2 :: Local Customizations ... SUCCESS [ 3.480 s]
[INFO] Apache Solr Webapp ... SUCCESS [ 54.792 s]
[INFO] DSpace SOLR :: Local Customizations ... SUCCESS [ 13.323 s]
[INFO] DSpace ORI-PMH ... SUCCESS [ 17.934 s]
[INFO] DSpace ORI-PMH :: Local Customizations ... SUCCESS [ 4.211 s]
[INFO] DSpace Assembly and Configuration ... SUCCESS [ 25.365 s]
[INFO] DSpace XML-UI Mirage2 Theme ... SUCCESS [ 0.300 s]
[INFO] BUILD SUCCESS
[INFO] Total time: 07:27 min
[INFO] Finished at: 2017-08-21T12:17:54-05:00
[INFO] Final Memory: 97M/748M
[INFO]
I:\install\dspace-5.2-src-release\dspace-5.2-src-release>
  
```

Figura 15. Configuración de DSpace (pantallas de instalación).

Diseño Técnico de los Datos

- Configuración de hoja de trabajo con campos, según normativa OpenAIRE y DataCite, con base en los Lineamientos Generales de Ciencia Abierta y los Lineamientos Específicos para Repositorios de Conacyt y según la naturaleza de los datos. Estos lineamientos pueden ser consultados en el

portal del Conacyt, en el apartado del *Programa de Repositorios* y que se actualizan constantemente.⁷⁴

- Registro e incrustación de los Metadatos que se emplearán en el marcaje de los datos y objetos digitales.

```

1 CAMPO: <field> </field>
2
3 ESQUEMA + ELEMENTO + CUALIFICADOR
4 <dc:schema dc </dc:schema>
5
6 <dc:elemento contribuidor </dc:elemento>
7 <dc-qualifier autor </dc-qualifier>
8
9 Repetente
10 <repeatable true </repeatable>
11 <repeatable false </repeatable>
12
13 Etiqueta
14 <label autor (s) </label>
15
16 ===== CELDAS DE INGRESO =====
17 <input-type name </input-type>
18 <input-type textarea </input-type>
19 <input-type fecha </input-type>
20 <input-type onebox </input-type>
21 <input-type value-pairs-name = "common_identifiers" qualrop_value </input-type>
22 <input-type value-pairs-name = "common_types" dropleft_wn </input-type>
23
24 AVISOS
25 <hint> Ingrese los apellidos y nombres del autor del artículo. </hint>
26
27
28 REQUERIDO
29 <required </required>
30
31 <required debe ingresar el dato del autor para pod e continuar </required>

```

Figura 16. Configuración de hoja de trabajo.

Figura 17. Hoja de trabajo con campos.

```

oai_dc:app.webui.json.SubmissionLookupRequest = submissionLookup, \
oai_dc:app.webui.json.UploadResponseForm = uploadResponseForm, \
oai_dc:app.webui.handle.handleSubmissionResponse = handleSubmissionResponse

### title - locale / language ###
# Default locale
# & locale in the form country or country_language or country_language_variant
# If no default locale is defined the server default locale will be used.
default.locale = es

### All the locales, that are supported by this instance of DSpace
# & comma-separated list of locales. All types of locale: country, country_language, country_language_variant
# Note that the appropriate file are present, especially that all the Message_x-properties are there
# May be used, e. g. webui.supported.locales = es, de

### Submission license substitution variables ###
# It is possible include contextual information in the submission license using substitution variables
# The text substitution is driven by a plugin implementation
plugin.name=oai_dc:app.content.License.LicenseArgumentFormater = \
oai_dc:app.content.License.SimpleDSpaceObjectLicenseFormater = collection, \
oai_dc:app.content.License.SimpleDSpaceObjectLicenseFormater = item, \
oai_dc:app.content.License.SimpleDSpaceObjectLicenseFormater = person

### Synchronization feed (RSS) Settings ###
# enable synchronization feeds - links display on community and collection home page
# (this setting is not used by RSSUI, so you enable feeds in your theme)
webui.feed.enable = true
# number of DSpace items per feed (the most recent submissions)
webui.feed.items = 5
# maximum number of feeds in memory cache
# value of 0 will disable caching
webui.feed.cache.size = 100
# number of hours to keep cached feeds before checking currency

```

Figura 18. Configuración de metadatos.

Figura 19. Visualización de los metadatos.⁷⁵

- Configuración de la contestación con base en su prefijo de metadatos para Datos en este caso Datacite.

```

rigeoterma.geofisica.unam.mx/oai/request?verb=ListRecords&metadataPrefix=oai_datacite

```

Figura 20. Prefijo de los metadatos metadataPrefix=oai_datacite.⁷⁶

Flujo de trabajo

- Establecimiento del flujo de trabajo (workflow) y revisión de los roles de reglas entre los administradores, revisores y publicadores.
- Creación de Comunidades. Se consideraron bajo la tipología de los datos que se englobarían en cada una de ellas.
 - Subcomunidades.

⁷⁴ México. Conacyt (2019). “Documentos de consulta de Política de Ciencia Abierta del Conacyt”. Disponible en <https://www.repositorionacionalcti.mx/documentos>. (Consultado: 25-06-2019).

⁷⁵ Las figuras 16-19 son tomadas del proceso de registro de los datos que se trabajan en DSpace. Pantallas obtenidas en el manejo del software.

⁷⁶ La figura 20 es tomada el URL del Repositorio de recursos geotérmicos para México.

- ii. Colecciones: el registro de cada uno de los datos y tipo de objeto que se desarrolla.
- iii. Ítems: el registro final del dato.

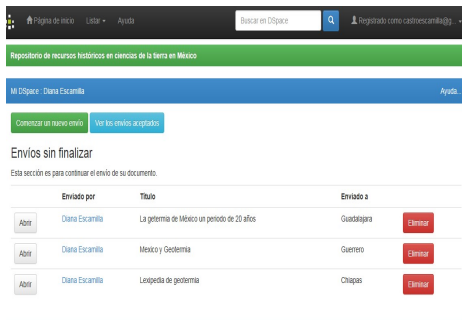


Figura 21. Roles de reglas entre los administradores, revisores y publicadores.

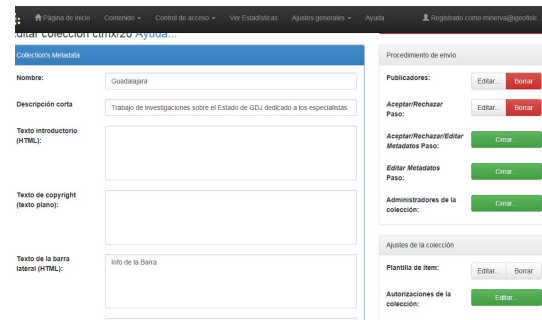


Figura 22. Creación de Comunidades, según tipología de los datos.

Estructura de despliegue de datos

- Configuración para visualización de nuevos campos en la interfaz del software externo para la consulta y gestión de los datos así como su visualización de ellos.
- Presentación de las miniaturas de portadas e imágenes. Esta configuración solo aplica para las publicaciones, videos, fotografías, mapas y todo aquel material documental que se quiera incluir.

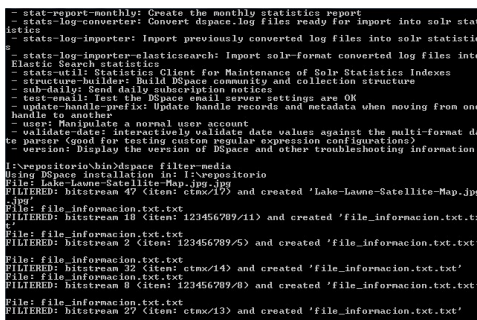


Figura 23. Configuración para visualización campos.

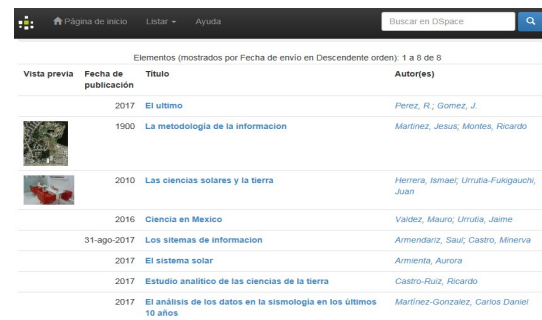


Figura 24. Visualización de las miniaturas de portadas e imágenes.

Interoperabilidad

- Configuración de estadísticas con base en las condiciones y lineamientos que solicita Conacyt para revisar la usabilidad de los recursos.
- Configuración de la lectura del Protocolo OAI desde interfaz pública y revisión de la salida de cosecha de datos.
- Diseño de plantilla para importar metadatos desde interfaz gráfica y configuración de la exportación de metadatos.

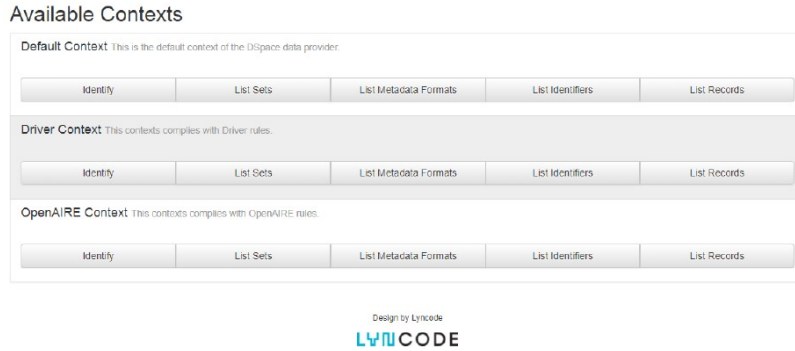


Figura 25. Protocolo OAI desde interfaz pública⁷⁷.

Diseño web y presentación final

- Personalización de interfaz gráfica JSPUI, con base en la funcionalidad de consulta e imagen institucional.
- Implementación de complementos como de Google Analytics, Statcounter, Awstats, según el más conveniente para complementar el conteo de uso de los datos o publicaciones.
- Gestión de copias de seguridad de bases de datos y archivos.

```

1832 # This feature is enabled by default to help prevent session hijacking
1833 # but may cause problems for shibboleth, etc
1834 #
1835 # webui.session.invalidate = true
1836
1837 # If you would like to use Google Analytics to track general website statistics then
1838 # use the following parameter to provide your Analytics key. First sign up for an
1839 # account at http://analytics.google.com, then create an entry for your repository
1840 # website. Analytics will give you a snippet of JavaScript code to place on your site,
1841 # inside that snippet is your Google Analytics key usually found in this line:
1842 # _uacct = "UA-XXXXXXX-X"
1843 # Take this key (just the UA-XXXXXXX-X part) and place it here in this parameter.
1844 jspui.google.analytics.key=UA-52588708-1
1845
1846 #-----#
1847 #-----XMLUI SPECIFIC CONFIGURATIONS-----#
1848 #-----#

```

Figura 26. Personalización de interface y complementos de Google Analytics.

3.4. Estructura

La estructura del Repositorio Institucional de Datos ofrece una serie de servicios bajo condiciones de Acceso Abierto y recursos de información de acuerdo con la naturaleza de los datos bajo una estructura de reutilización mediante estándares internacionales e interoperables que, además preserva, disemina e intercambia datos y los comparte con otros pares bajo un ambiente de colaboración y creación de otras investigaciones.

⁷⁷ UNAM. SSN (2019). "Pantalla del protocolo OAI-PMH del Repositorio Institucional de Datos del Servicio Sismológico Nacional". Disponible en <http://rigeofisica.ssn.unam.mx/jspui/>. (Consultado: 18-06-2019).

A continuación se presenta la estructura Repositorio Institucional de Datos desde la interoperabilidad con el Repositorio Nacional:



Figura 27. Estructura del Repositorio Institucional y su relación con el Repositorio Nacional.

En cuanto a la organización y el personal de trabajo que participa en el desarrollo del proyecto, se emprende el siguiente orden en cual intervienen un grupo de especialistas en diversas áreas del conocimiento para poder plasmar diferentes contextos en el momento de la toma de decisiones en el diseño del Repositorio. Esta es la estructura que se tomó en la BCCT para el desarrollo de los RI que se trabajaron y que, bajo una evaluación con el Conacyt, es ampliamente funcional para otros proyectos similares:

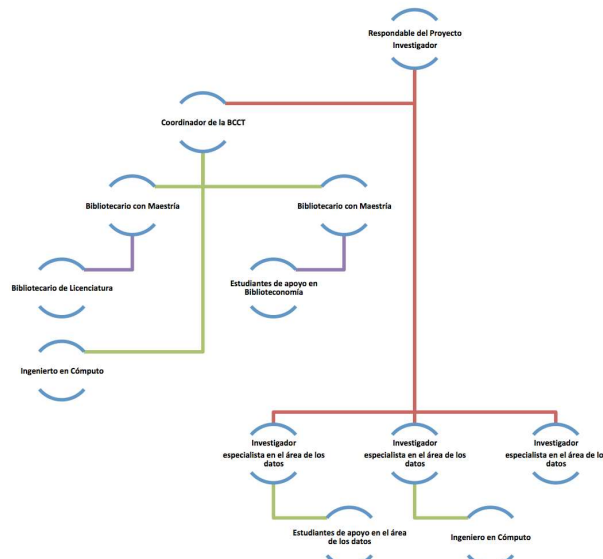


Figura 28. Organigrama de trabajo para el Repositorio de Datos.

Cada definición de los roles de los participantes dentro de la Plataforma de DSpace se estructura en 4 grupos, como se presenta en la figura 28, con el fin de trabajar mediante una serie de autorizaciones que requieren configuración, así

como la asignación de los diferentes niveles que se ejecutarán en el registro del contenido en cada una de las comunidades correspondientes, siendo estos:



Figura 29. Organización de los grupos para desarrollar las actividades.

La designación de las funciones del personal del proyecto se definieron de la siguiente manera:

- a) *Responsable del Proyecto*: permiso de administrador y coordinador del proyecto del RI y el enlace con los organismos que participarán.
- b) *Publicadores y Editores*: 1 bibliotecario con Maestría en Bibliotecología y 1 investigador especialista de los datos.
- c) *Estudiantes de Posgrado*: organización, revisión y asignación de nombres de los archivos en texto completo que se trabajarán para la inclusión.
- d) *Revisor*: 1 bibliotecario con Maestría en Bibliotecología y 1 de Licenciatura.
- e) *Modificación de datos*: 1 bibliotecario con Maestría en Bibliotecología y 1 investigador especialista en los datos.

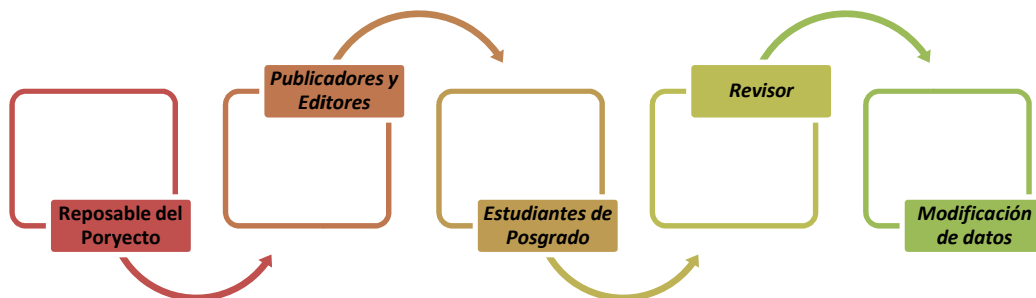


Figura 30. Roles de los participantes dentro de la plataforma de DSpace.



Figura 31. Presentación de la autorización y revisión del trabajo de los participantes dentro de la Plataforma de DSpace.⁷⁸

En ese sentido, es importante mencionar que en los grupos de trabajo laboran bibliotecarios, especialistas en los diferentes datos en Ciencias de la Tierra y personal de posgrado especializados en el área, a fin de llevar a cabo el desarrollo de las interfaces, así como con los parámetros correspondientes para la conectividad, recuperación y visualización de los datos, mediante las estrategias más adecuadas que permitan interconectar las estructuras del registro y del dato.

Los avances se programaron para su presentación cada bimestre mediante un informe de actividades de cada uno de los integrantes según su roles y responsabilidades designados. Los canales de información son a través de correo electrónico, vía Skype y reuniones de trabajo en la sala de juntas de la BCCT y a través de la nube.

Indicadores de evaluación

Para el desarrollo del Repositorio se tomaron los siguientes indicadores, a fin de poder evaluar la progresión, medición, metas y objetivos, dividiéndolos en dos líneas de acción:

Independencia

- Administración
 - Productividad: se revisará semanalmente la cantidad de registros que se ingresen al repositorio, así como también se medirá el promedio de datos generados por los sistemas.
 - Calidad de la información: verificar la calidad de los datos y el formato de los objetos digitales que se hospedarán, por medio de los editores y revisores en DSpace.

⁷⁸ UNAM. IGEF. “Roles de autorización y revisión que se utilizan en el Repositorio de Recursos Geotérmicos para México”. Disponible en <http://rigeotermia.geofisica.unam.mx/jspui>. (Consultado: 25-06-2019).

Verificabilidad

Mediante el módulo de estadísticas se realizará un registro de la producción y actividad concluida en el Repositorio a través de la:

- Producción: cantidad de registros que ingresan y comparación de forma mensual.
- Medición:
 - Consulta: revisar qué tipo de comunidades, subcomunidades y colecciones se accede más.
 - Uso: saber que se consulta, quién lo solicita y si se puede conocer en qué lo aplican.
 - Descarga: medir como resultado qué es lo que descargan como producto final.
- Accesibilidad: con base en la estructura que se tiene en el RI, compararlo con los demás repositorios y realizar pruebas de cosecha con el OAI.
- Establecer colaboración con otras dependencias con quién se pueda compartir la información y poder enlazar entre repositorios.

3.5 Comunidades

En cuanto a la creación y la política de las comunidades, subcomunidades y colecciones, el énfasis es albergar conjunto de datos según su naturaleza y área temática, pero también se toma en cuenta la producción de los datos y el material científico –como artículos, presentaciones de congresos, informes técnicos, audiovisuales, imágenes y mapas–.

Es necesario pensar en una organización correcta, debido a que existen diversas formas en el universo de datos o de información. Aquí mostramos un ejemplo de cómo se realiza en DSpace:



Figura 32. Nivel jerárquico entre la comunidad, subcomunidad, colección y el ítem.

En nuestro caso, se determinaron con base en una serie de elementos entendibles, que englobó la tipología de los datos. Estos fueron:

- Tipificación de los datos.
- Organización con base en su naturaleza (sismogramas, geotermias, químicas, clima espacial, temperatura, profundidad, entre otros).
- Cotejo de los campos para posteriormente asignarlos, ya sea la subcomunidad o colección según corresponda.
- Revisión de los recursos con base en un formato y extensión.

Con el fin de poder contar con un contexto más amplio de cómo podemos organizar los datos, presentamos una revisión y comparación que se realizó en 96 Repositorios Institucionales de Datos y Literatura ambos tipos interoperando con el Repositorio Nacional en México, para lo cual es necesario **consultar el Anexo 1** al final del trabajo, en donde se presentan diversas características de estos dos tipos de Repositorios.

Con base en la revisión y el análisis los resultados son los siguientes en cuanto a la creación de comunidades: el 87% de los Repositorios que interoperan con el Repositorio Nacional que incluyen Literatura en su contenido sus comunidades las estructuran en Formatos, Áreas académicas, Temáticas y Administrativas.

Estructura	Asignación de nombres en Comunidades de los RI de Literatura								
Formatos	Tesis	Libros	Artículo	Capítulos Libros	Memorias Congreso	Poster	Multimedia	Patentes	
Académicas	Docencia	Investigación	Producción Científica	Patrimonio	Editorial	Laboratorio de Investigación	Academia	Divulgación	
Temáticas	Aeronáutica	Ingeniería	Ciencias Sociales	Ciencias Naturales	Salud	Arte	Derecho	Literatura	
Administrativas	Docentes	Proyectos	Casos	Informes	Manuales	Trabajo en general	Reportes		

El 13% que refiere a los Repositorio de Datos y el objeto central de ésta investigación su creación de comunidades las estructuran por Áreas Geográficas, Coordenadas, Manifestaciones, Tipologías, Taxonomías, por Tiempos y Espacio entre otras, debido a que existe una amplia y variada gama de datos que son generados por las Entidades

ESTRUCTURA	Asignación de nombres en Comunidades de los RI de Datos								
GEOGRAFICAS	Estados	Zonas	Municipios	Forma	Latitud Altitud	Ubicación GPS	Coordenadas	Manifestación	Especificidad del dato
TIPOLOGÍA	Familia	Taxonomía	Espacio	Clima	Ambiente	Tiempo	Velocidad	Medición	ADN

Tablas que presentan el tipo de Comunidades que se diseñaron en cada uno de los 96 RI tanto de Datos como de Literatura de las instituciones públicas y privadas (datos actualizados al 30 de junio de 2019).

Como se puede observar, las comunidades están creadas con base en la estructura del formato, áreas académicas, temáticas, administrativas, geográficas y por tipología. De cada una de ella se desprenden subcomunidades o colecciones, según las necesidades de la institución autora del RI y, sobre todo, en la organización de los datos u objetos digitales.

Es importante remarcar que las Comunidades de los RI que están interoperando con el Nacional son de literatura y datos, y en ambos casos se manejan objetos digitales –como set de datos y publicaciones científicas, patentes, tesis, información científica y tecnología dirigida a alumnos, profesores, público en general e investigadores, entre otros documentos y usuarios, como se pueden apreciar en el **Anexo 1**–. Cada repositorio establece a quién va dirigida la información que publica, así como la disponibilidad del acceso del texto completo.

Para contar con una visión más amplia de lo incluido en los RI, presentamos a continuación dos ejemplos de Comunidades de Repositorio de Datos:

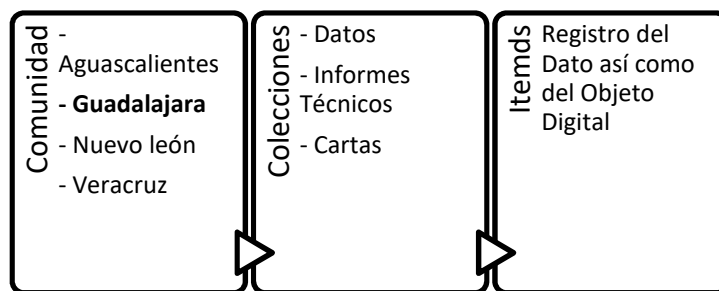


Figura 33. Ejemplo de las comunidades del Repositorio Institucional de Recursos Geotérmicos para México.

Envío: elegir colección [Más ayuda...](#)

Seleccione la colección a la que desea enviar un nuevo ítem, y haga clic en "Siguiente".

Colección

[Ir a](#)
[Página de Inicio](#)
[Mi DSpace](#)

Figura 34. Registro de la Comunidad de Guadalajara (Comunidades geográficas) en el Repositorio Institucional de Recursos Geotérmicos para México.⁷⁹

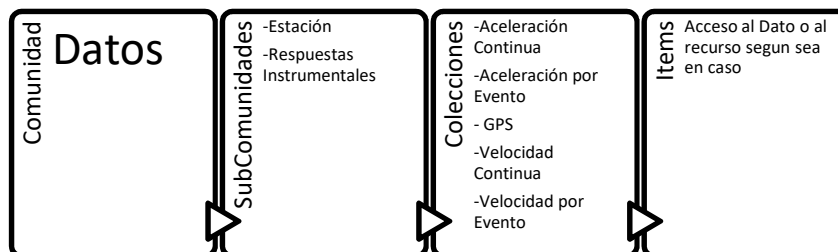


Figura 35. Comunidades del Repositorio Institucional del SSN de México.

3.6 Metadatos

La descripción de los Datos se realiza es mediante metadatos, esto es, información estructurada que facilita su recuperación, acceso, visibilidad, así como su preservación. Además de mostrar una descripción precisa y completa que proporciona el contexto a los Datos, permitiendo que puedan ser encontrados, acceder a ellos, interoperables y, sobre todo, ser reutilizados.

⁷⁹ Op. Cit.

Las funciones que se implementan con los metadatos son:

- Descubrimiento, cualquier buscador utiliza texto en sus búsquedas, no serían recuperables a no ser por el metadato.
- Recuperación de los metadatos facilita la búsqueda por la descripción del dato o documento.
- Localización por medio de identificadores a una entidad de forma única, estable, accesible y persistente.
- Utilización en el apoyo del uso de un conjunto de datos. Los investigadores necesitan comprender los datos: estructura, definiciones de términos utilizados, cómo se han recopilado y cómo se deben leer.
- Reutilización, los investigadores requieren reutilizar datos ya existentes para un nuevo proyecto investigación. Es importante una interpretación para saber si son fiables.
- Preservación, en este punto los metadatos son primordiales para conocer si el recurso está íntegro o ha sufrido transformaciones.

Los Datos que se incluyen en RI se analizaron y se marcaron bajo la actualización de los Lineamientos Específicos para Repositorios del Apéndice 2. Metadatos para datos para los repositorios institucionales ver. 3.0.⁸⁰

1. Identificador

Núm.	Campo LER	Campo OpenAIRE	Uso OpenAIRE	Uso LER	Comentarios
1	Identificador del recurso	Identifier	Obligatorio	Obligatorio	Cadena única que identifica un recurso. El formato debe ser "10.1234/foo" Lista controlada de valores (de acuerdo con OpenAIRE)
1,1	Tipo de identificador	identifierType	Obligatorio	Obligatorio	DOI Handle ARK PURL URN URL
Ejemplo		<identifier identifierType="Handle">http://rigeofisica.ssn.unam.mx/jspui/handle/ssnmx/59</identifier>			

2. Autor

Núm.	Campo LER	Campo OpenAIRE	Uso OpenAIRE	Uso LER	Comentarios
2	Autor	Creator	Obligatorio	Obligatorio	Los principales investigadores involucrados en la producción de los datos o de la publicación, en orden de prioridad. El autor del recurso puede ser una persona, una organización o una

⁸⁰ México. Conacyt (2019). "Lineamientos Específicos para Repositorios del Apéndice 2. Metadatos para datos para los Repositorios Institucionales 3.0". Disponible en <http://www.siicyt.gob.mx/index.php/normatividad/2-conacyt/1-programas-vigentes-normatividad/lineamientos/lineamientos-especificos-para-repositorios/4704-lineamientos-especificos-para-repositorios/file>. (Consultado: 30-junio-2019).

					institución.
2,1	Nombre de autor	creatorName	Obligatorio	Obligatorio	El autor del recurso puede ser una persona o una organización o una institución. Se consumirá de manera automática el catálogo de autores de Conacyt disponible en formato REST. Deberán presentarse en el siguiente formato: Nombre(s), Primer Apellido Segundo Apellido. Si el metadato se halla ausente, el registro será rechazado.
2,2	Identificador de autor	nameIdentifier	Recomendado	Obligatorio	Este elemento tomará un atributo obligatorio llamado ID, que deberá contener un identificador único. Si el metadato se halla ausente, el registro será rechazado. Se deberá contar con al menos un autor con identificador válido.
2.2.1	Esquema de identificador de autor	nameIdentifierScheme	Recomendado	Obligatorio	Los identificadores a utilizar serán CURP, CVU, ORCID y DNI.
2.2.2	URI del esquema de identificador de autor	schemeURI	Recomendado	Obligatorio	URI del esquema de identificador de autor a utilizar.
2,3	Afiliación del autor	affiliation	Recomendado	Recomendado	Es la afiliación organizacional o institucional del autor.
Ejemplo	<pre> <creators> <creator> <creatorName>Servicio Sismológico Nacional (IGEF-UNAM)</creatorName> <affiliation>SSN, IGEF-UNAM </affiliation> <nameIdentifier nameIdentifierScheme="RNCTIMX" schemeURI="http://repositorionacionalcti.mx/"> info:eu-repo/dai/mx/rn/3190557 </nameIdentifier> </creator> </creators> </pre>				

3. Título

Núm.	Campo LER	Campo OpenAIRE	Uso OpenAIRE	Uso LER	Comentarios
3	Título	Title	Obligatorio	Obligatorio	Se refiere al título y a cualquier subtítulo por el que el recurso es conocido. Si el metadato se halla ausente, el registro es rechazado.
3,1	Tipo de título	titleType	Opcional	Opcional	Lista controlada de valores (de acuerdo con OpenAIRE) Alternative Title (Título Alternativo) Subtitle (Subtítulo) TranslatedTitle (Título Traducido)

Ejemplo	<pre><titles> <title> Datos de velocidad en continuo a 100 muestras por segundo</title> <title titleType="Subtitle">No aplica</title> </titles></pre>
---------	---

4. Editor

Núm.	Campo LER	Campo OpenAIRE	Uso OpenAIRE	Uso LER	Comentarios
4	Editor	Publisher	Obligatorio	Obligatorio	Nombre de la entidad, típicamente un editor, responsable de hacer que la versión del recurso esté disponible. Una persona, una organización o un servicio pueden ser un editor. De ser posible, se utilizará una lista controlada (de editores, títulos de journals e ISSN) como la base de datos SHERPA/RoMEO.1
Ejemplo		<publisher>UNAM, IGEF, SSN, Grupo de Trabajo</publisher>			

5. Año de Publicación

Núm.	Campo LER	Campo OpenAIRE	Uso OpenAIRE	Uso LER	Comentarios
5	Año de Publicación	PublicationYear	Obligatorio	Obligatorio	Año en que los datos fueron o serán puestos a disposición pública. Formato: YYYY. Si los datos se encuentran bajo un período de embargo, colocar el año en que el período de embargo finalizará. Véase 8. Fecha y 8.1 Tipo de Fecha para la correcta aplicación del período de embargo (si aplica).
Ejemplo		<publicationYear>2007</publicationYear>			

6. Materia

Núm.	Campo LER	Campo OpenAIRE	Uso OpenAIRE	Uso LER	Comentarios
6	Materia	Subject	Recomendado	Obligatorio cuando aplique	El área de conocimiento del Conacyt es obligatoria en su primer nivel. Debe ser asignada consumiendo el catálogo para este fin. Adicionalmente, y con la intención de agregar detalle, se podrán utilizar palabras clave, descriptores, frases y/o códigos de clasificación utilizados para describir la materia del recurso.

6,1	Esquema de materia	subjectScheme	Opcional	Obligatorio cuando aplique	Deberá utilizarse la estructura establecida en OpenAIRE 2.0 para los casos en que se especifique la fuente.
6,2	URI del esquema de materia	schemeURI	Opcional	Opcional	URI del esquema de materia (independiente al Catálogo de Áreas del Conocimiento de Conacyt) que se utilizará. No se usa para áreas del conocimiento (CTI) Ejemplo: http://id.loc.gov/authorities/subjects http://dewey.info/
Ejemplo	<pre> <subjects> <subject subjectScheme="LCC" schemeURI="http://id.loc.gov/authorities/subjects/sh85119634">Sismologia </subject> <subject subjectScheme="GeoRef" schemeURI="https://pubs.geoscienceworld.org/georef/term/5/30320/seismology">Sismología </subject> </subjects> <subjects> <subject subjectScheme="CTI">1</subject> </subjects> <subjects> <subject subjectScheme="CTI">25</subject> </subjects> <subjects> <subject subjectScheme="CTI">2507</subject> </subjects> <subjects> <subject subjectScheme="CTI">250705</subject> </subjects> </pre>				

7. Contribuidor

Núm.	Campo LER	Campo OpenAIRE	Uso OpenAIRE	Uso LER	Comentarios
7	Colaborador (es) u otros Autor(es)	Contributor	Obligatorio cuando aplique	Obligatorio cuando aplique	Institución o persona responsable de recolectar, administrar, distribuir o contribuir de otra manera al desarrollo del recurso.
7,1	Tipo de colaborador	contributorType	Obligatorio cuando aplique	Obligatorio cuando aplique	Lista controlada de valores: ContactPerson DataCollector DataCurator DataManager Distributor Editor Funder HostingInstitution Producer ProjectLeader ProjectManager ProjectMember RegistrationAgency RegistrationAuthority RelatedPerson

					Researcher ResearchGroup RightsHolder Sponsor Supervisor WorkingPackageLeader Other
7,2	Nombre de colaborador	contributorName	Obligatorio cuando aplique	Obligatorio cuando aplique	Si se utiliza el elemento 7. Colaborador(es) u otros autor(es) es obligatorio. El nombre deberá ser desplegado en el formato Nombre(s), Primer Apellido, Segundo Apellido.
7,3	Identificador de colaborador	nameIdentifier	Obligatorio cuando aplique	Obligatorio cuando aplique	Persona, organización o servicio responsable por haber hecho contribuciones al recurso. Este elemento tomará un atributo llamado ID, que deberá contener un identificador único HTTP URI para el colaborador. Se añadirá un elemento por cada colaborador, en orden de significancia, con el respectivo formato para persona o institución: "http://orcid.org/0000-0000-0000-0000">Nombre(s) Apellidos
7.3.1	Esquema de identificador de colaborador	nameIdentifierScheme	Obligatorio cuando aplique	Obligatorio cuando aplique	Los identificadores a utilizar serán CURP, CVU, ORCID y DNI.
7.3.2	URI del esquema de identificador de colaborador	schemeURI	Opcional	Opcional	URI del esquema de identificador de autor a utilizar. Por ejemplo: http://orcid.org
7,4	Afiliación de colaborador	affiliation	Opcional	Opcional	Es la afiliación organizacional o institucional del colaborador.
Ejemplo	<pre> <contributors> <contributor contributorType="Producer"> <contributorName>Servicio Sismológico Nacional, Red Nacional de Observatorios Sismológicos, (MX), Tuxtla Gutiérrez, Chis., TGIG</contributorName> <nameIdentifier nameIdentifierScheme="RNCTIMX" schemeURI="http://repositorionacionalcti.mx/"> info:eu-repo/dai/mx/rn/3190483 </nameIdentifier> </contributor> <contributor contributorType="ProjectLeader"> <contributorName>NOMBRE DE LA PERSONA</contributorName> <nameIdentifier nameIdentifierScheme="RNCTIMX" schemeURI="http://repositorionacionalcti.mx/"> info:eu-repo/dai/mx/orcid/0000-0000-0000-0000 </nameIdentifier> </contributor> <contributor contributorType="ProjectMember"> <contributorName>NOMBRE DE LA PERSONA </contributorName> <nameIdentifier nameIdentifierScheme="RNCTIMX" schemeURI="http://repositorionacionalcti.mx/"> info:eu-repo/dai/mx/orcid/0000-0000-0000-0000-</nameIdentifier> </contributor> </contributors> </pre>				

8. Fecha

Núm.	Campo LER	Campo OpenAIRE	Uso OpenAIRE	Uso LER	Comentarios
8	Fecha	Date	Obligatorio	Obligatorio	Fechas relevantes al trabajo. Los formatos aceptados son YYYY, YYYY-MM-DD, YYYY-MM-DDThh : mm : ssTZD o cualquier otro formato o nivel de granularidad descrito en W3CDTF (ISO 8601).
8,1	Tipo de fecha	dateType	Obligatorio	Obligatorio	<p>Lista controlada de valores:</p> <ul style="list-style-type: none"> Accepted Available Copyrighted Collected Created Issued Submitted Updated Valid <p>Nota: Se utiliza "Issued" para la fecha en que el recurso ha sido publicado o distribuido. Para indicar la fecha de inicio de un embargo, utilizar "Accepted". Para indicar la finalización de un embargo, utilizar "Available".</p>
Ejemplo		<pre><dates> <date dateType="Issued">2005-04-05</date> </dates> Embargo: <dates> <date dateType="Accepted">2011-12-01</date> <date dateType="Available">2012-12-01</date> </dates></pre>			

9. Idioma

Núm.	Campo LER	Campo OpenAIRE	Uso OpenAIRE	Uso LER	Comentarios
9	Idioma	Language	Recomendado	Recomendado	Idioma principal del recurso de información. Se utilizarán valores de acuerdo a ISO 639-3
Ejemplo		<pre><language>spa</language></pre>			

10. Tipo de recursos

Núm.	Campo LER	Campo OpenAIRE	Uso OpenAIRE	Uso LER	Comentarios
10	Tipo de Recurso	ResourceType	Recomendado	Obligatorio	Descripción del recurso de información que se presenta. Se inicia con el ResourceType y se complementa con el resourceTypeGeneral. Consultar la lista

					controlada de valores (10.1 resourceTypeGeneral) para su implementación
10,1	Tipo general del recurso	resourceTypeGeneral	Recomendado	Obligatorio	Lista controlada de valores: Audiovisual Collection Dataset Event Image InteractiveResource Model PhysicalObject Service Software Sound Text Workflow
Ejemplo		<code><resourceType resourceTypeGeneral="Dataset">Conjunto de datos sísmológicos</resourceType</code>			

11. Identificador alternativo

Núm.	Campo LER	Campo OpenAIRE	Uso OpenAIRE	Uso LER	Comentarios
11	Identificador Alternativo	AlternateIdentifier	Opcional	Opcional	Identificador o identificadores correspondientes al recurso, distintos al identificador principal aplicado al recurso.
11,1	Tipo de identificador alternativo	alternateIdentifierType	Opcional	Obligatorio cuando aplique	Describe el tipo de identificador alternativo. Si el elemento 11. Identificador Alternativo es utilizado, se debe utilizar este metadato de forma obligatoria.
Ejemplo		<pre><alternateIdentifiers> <alternateIdentifier alternateIdentifierType="ISBN"> 937-0-1234-56789-X </alternateIdentifier> </alternateIdentifiers></pre> NO APLICA			

12. Identificador relacionado

Núm.	Campo LER	Campo OpenAIRE	Uso OpenAIRE	Uso LER	Comentarios
12	Identificador Relacionado	RelatedIdentifier	Obligatorio cuando aplique	Obligatorio cuando aplique	Identificador o identificadores relacionados con el recurso. Deberán ser identificadores únicos y estandarizados.
12,1	Tipo de identificador relacionado	relatedIdentifierType	Obligatorio	Obligatorio	Lista controlada de valores: ARK arXiv bibcode DOI EAN13 EISSN Handle ISBN ISSN ISTC LISSN LSID PMID PURL UPC URL

					URN
12,2	Tipo de relación	relationType	Obligatorio cuando aplique <i>Se emplea cuando el recurso registrado tiene relación con otro producto y/o publicación u objeto digital, etc.</i>	Obligatorio	<p>Describe la relación entre el recurso que está siendo registrado (A) y el recurso relacionado (B). Lista controlada de valores:</p> <ul style="list-style-type: none"> * IsCitedBy (indica que B incluye a A en una cita) * Cites (indica que A incluye a B en una cita) * IsSupplementTo (indica que A es un suplemento de B) * ISupplementedBy (indica que B es un suplemento de A) * IsContinuedBy (indica que A continúa con el trabajo B) * Continues (indica que A es una continuación del trabajo B) * HasMetadata (indica que el recurso A tiene metadatos adicionales de B) * IsMetadataFor (indica metadatos adicionales de A para el recurso B) * IsNewVersionOf (indica que A es una nueva versión de B, donde la nueva edición ha sido modificada o actualizada) * IsPreviousVersionOf (indica que A es una edición previa de B) * IsPartOf (Indica que el recurso A es una porción de B; puede ser utilizado como elementos de una serie) * HasPart (indica que A incluye parte de B) * IsReferencedBy (indica que A es utilizado como un recurso de información por B) * References (indica que B es utilizado como un recurso de información por A) * IsDocumentedBy (indica que B es documentación sobre o que explica a A) * Documents (indica que A es un documento sobre o que explica a B) * isCompiledBy (indica que B es usado para compilar o crear A) * Compiles (indica que B es el resultado de una compilación o creación al utilizar A) * IsVariantFormOf (indica que A es una variante o forma diferente de B) * IsOriginalFormOf (indica que A es la forma original de B) * IsIdenticalTo (indica que A es idéntico a B, se utiliza cuando es necesario registrar dos instancias separadas de un mismo evento) * IsReviewedBy (indica que A es revisado por B) * Reviews (indica que A es una revisión de B) * IsDerivedFrom (indica que B es un

					recurso basado en A) * IsSourceOf (indica que A es un recurso basado en B)
12,3	Esquema de metadatos relacionado	relatedMetadataScheme	<pre><description> <description descriptionType="Methods">Los datos de velocidad de la Red Sismológica del SSN se obtienen con sismómetros triaxiales de banda ancha con respuesta plana a la velocidad del suelo entre 0.01 a 30 Hz. Se graban a 100 muestras por segundo, son convertidas a un formato digital en el registrador y transmitidos vía remota, mediante satélite o internet, al centro de monitoreo donde son recibidos de forma continua y almacenados.</description> </description></pre>	Opcional	Se utiliza únicamente cuando el tipo de relación es HasMetadata o IsMetadataFor
12.3.1	URI del esquema de identificador relacionado	schemeURI	Opcional	Opcional	Se utiliza únicamente cuando el tipo de relación es HasMetadata o IsMetadataFor

12.3.2	Tipo de esquema de metadatos relacionado	schemeType	Opcional	Opcional	Indica el tipo de relación con el esquema de metadatos. Se utiliza únicamente cuando el tipo de relación es HasMetadata o IsMetadataFor. Ejemplos: XSD, DDT, Turtle.
Ejemplo		<code><relatedIdentifier relatedIdentifierType="DOI" relationType="IsSupplementTo">10.21766/SSNMX/SN/MX</relatedIdentifier></code>			

13. Tamaño

Núm.	Campo LER	Campo OpenAIRE	Uso OpenAIRE	Uso LER	Comentarios
13	Tamaño	Size	Opcional	Opcional	Tamaño del recurso. Por ejemplo: 15 páginas, 6 MB <code><sizes></code> <code><size>15 pages</size></code> <code><size>6 MB</size></code> <code></sizes></code>
Ejemplo		<code><sizes> <size>21 temp.</size> </sizes></code>			

14. Formato

Núm.	Campo LER	Campo OpenAIRE	Uso OpenAIRE	Uso LER	Comentarios
14	Formato	Format	Opcional	Opcional	Se refiere al formato del recurso. Ejemplo: PDF, XML, etcétera. Véase http://www.iana.org/assignments/media-types/media-types.xhtml <code><formats></code> <code><format>PDF</format></code> <code><format>application/pdf</format></code> <code></formats></code>
Ejemplo		<code><format>gz</format></code> <code><format>application/xml</format></code>			

15. Versión

Núm.	Campo LER	Campo OpenAIRE	Uso OpenAIRE	Uso LER	Comentarios
15	Versión	Version	Opcional	Opcional	Indica el número de versión del recurso. Puede ser utilizado en conjunto con las propiedades de 11. Identificador alternativo y 12. Identificador relacionado.
Ejemplo		<code><version>1.0</version></code>			

16. Nivel de acceso y condiciones de licencia

Núm.	Campo LER	Campo OpenAIRE	Uso OpenAIRE	Uso LER	Comentarios
16	Nivel de Acceso y condiciones de licencia.	Rights	Obligatorio cuando aplique	Obligatorio	Los derechos o condiciones de licencia a los que está sujeto el recurso. En la primera ocurrencia debe colocarse el Nivel de acceso de acuerdo con el catálogo establecido. En la segunda ocurrencia deben colocarse las condiciones de licencia de acuerdo con Creative Commons.
16,1	Condiciones de Licencia (RRF= Sugerencia de cambio de nombre del atributo: URI de nivel de acceso y condiciones de licencia)	rightsURI	Obligatorio cuando aplique	Obligatorio	Se coloca la URL o URI de la licencia seleccionada de acuerdo con los niveles de acceso de Creative Commons, consumiendo el catálogo de Licencia y Nivel de Acceso del Repositorio Nacional. Disponible en http://catalogs.repositorionacionalcti.mx/ Nota: El Repositorio Nacional no cosecha los recursos cuyo nivel de acceso sea "Restringido" pues no es su finalidad.
Ejemplo		<pre><rightsList> <rights rightsURI="info:eu-repo/semantics/openAccess" /> <rights rightsURI="http://creativecommons.org/licenses/by-nc-nd/4.0/"> Creative Commons Attribution 4.0 International </rights> </rightsList></pre>			

17. Descripción

Núm.	Campo LER	Campo OpenAIRE	Uso OpenAIRE	Uso LER	Comentarios
17	Descripción	Description	Obligatorio cuando aplique	Obligatorio cuando aplique	La descripción del contenido del recurso, usando texto libre, permite aplicar técnicas de minería de textos gracias a su indexación. Cuando sea posible, se recomiendan versiones en español, portugués e inglés para mayor visibilidad. Si la metodología y la descripción no son incluidas como archivos, deben estar incluidas en este elemento.
17,1	Tipo de descripción	descriptionType	Obligatorio cuando aplique *** Son los valores que más se utilizaron con el fin de explicar el Método por el cual se recoge o se crea el dato	Obligatorio cuando aplique	Lista controlada de valores: Abstract*** Methods*** SeriesInformation TableOfContents Other*** Ejemplo: <descriptions> <description descriptionType="Abstract"> Este es un resumen </description>

			y/o Evento.	<description descriptionType="Other"> Este es un resumen </description> </descriptions>
Ejemplo	<descriptions> <description descriptionType="Methods">Los datos de velocidad de la Red Sismológica del SSN se obtienen con sismómetros triaxiales de banda ancha con respuesta plana a la velocidad del suelo entre 0.01 a 30 Hz. Se graban a 100 muestras por segundo, son convertidas a un formato digital en el registrador y transmitidos vía remota, mediante satélite o internet, al centro de monitoreo donde son recibidos de forma continua y almacenados.</description> </descriptions>			

18. Geo-localización

Núm.	Campo LER	Campo OpenAIRE	Uso OpenAIRE	Uso LER	Comentarios
18	Geo-Localización	GeoLocation	Opcional	Opcional	Indica la ubicación espacial o nombre del lugar donde los datos fueron recolectados o al que los datos se refieren. Pueden indicarse tantas localizaciones como sean necesarias. Si se utiliza este elemento, debe tener, por lo menos, uno de los atributos listados bajo este elemento.
18,1	Punto de geo-localización	geoLocationPoint	Opcional	Opcional	Contiene un par de latitud-longitud, separados por un espacio en blanco.
18,2	Caja de geo-localización	geoLocationBox	Opcional	Opcional	Indica los límites espaciales de un lugar. Una caja contiene dos pares de latitud-longitud, separadas por espacios en blanco. Cada par debe ser separado a su vez por un espacio en blanco. El primer par es la esquina inferior (normalmente suroeste), mientras que el segundo par indica la esquina superior (normalmente noreste).
18,3	Lugar de geo-localización	geoLocationPlace	Opcional	Opcional	Texto libre. Se utiliza para describir una localización geográfica.
		<geoLocations> <geoLocation> <geoLocationPoint>16.77821 -93.122278</geoLocationPoint> <geoLocationBox>16.728 -93.172 16.828 -93.072</geoLocationBox> <geoLocationPlace>Chiapas, Tuxtla Gutierrez, "Albores Guillen, pomarrosa"</geoLocationPlace> </geoLocation> </geoLocations>			

Las actividades adicionales que conllevó la asignación de los metadatos, en los Repositorios Institucionales de Datos, fue según su especialidad y marcaje, según su propiedad por evento, recolección, fenómeno, temperatura, entre otros:

- El Repositorio de datos se componen de varios archivos en diversos formatos, es decir, la relación entre el Dato que está siendo registrado, en ocasiones por nacimiento o naturaleza se relaciona con otro recurso, por lo que hay que llevar a cabo su asociación. Es por eso que se utilizaron

metadatos asociados correspondientes en algunos casos con Datos, con documentos en formato PDF –como cartas hidrogeológicas, sismogramas, diagramas de climas, informes técnicos, resumen de congreso, artículos y, entre otros–, a los cuales se realizó un análisis para la elaboración de la citación y una catalogación bajo la norma APA, a fin de incluir los datos en las etiquetas correspondientes: *relatedIdentifier relatedIdentifierType=relationType* con valores de relación como: **IsSupplementTo, *IsSourceOf, *isCompiledBy, *IsPartOf, *IsCitedBy*.

- En cuanto a la asignación de temas en la especialidad de Ciencias de la Tierra, nos basamos en los de la Library of Congress (LCC), los de GeoRefs –que pertenecen a la Biblioteca del Congreso de Estados Unidos, al Sistema Internacional GeoScienceWorld, y los de la UNESCO, con el fin de lograr una localización más expedita y oportuna. Existe una variedad según la temática del Repositorio de datos para revisar los diferentes temas, vocabularios y encabezamientos de materia que utilizan y se aplican en los diferentes Repositorios Institucionales que actualmente interoperan con el Repositorio Nacional (consulte el **Anexo 1**) en donde se muestra los más utilizados y cómo se aplican según la especialidad.
- Es importante mencionar que se trabaja bajo software libre de DSpace por lo que fue necesario una configuración adicional para que contestara bajo ambiente DataCite, correspondiente a la salida de los Datos bajo el *metadata Prefix=oai_datacite*, y con ello realice una verificación constante de los metadatos.

A continuación se muestra la interface que tiene DSpace para el registro de campos de metadatos.

Figura 36. Registro de campos de metadatos.

Figura 37. Pantalla para añadir más campos de metadatos.⁸¹

⁸¹ UNAM. IGEF. Las figuras 36 y 37 fueron tomadas del Repositorio de Recursos Geotérmicos para México. *Op. Cit.*

3.7. Software de interface: características de enlace con el dato

Crear un Repositorio de Datos implica desarrollar una infraestructura técnica que apoye la accesibilidad en múltiples formatos y la preservación, por lo que es necesario buscar la asociación del enlace hacia el acceso a los datos y sí es el caso al recurso relacionado, así como su presentación y despliegue de los mismos –ya sea que se necesiten mostrar tablas o archivos con diversas extensiones como *.tar*, *.gz*, *.tgz*, *.gz* (archivos comprimidos), archivo *.fit* (se utilizan para almacenar datos relacionados con la astronomía que se capturan a través del telescopio), de texto como *.csv*, *.xls*, *.rtf*, *.figuras*, *.jpg*, *.gif*, *.png*, *.bmp*, gráficos, videos, animaciones *.mpg*, *.avi*, *.3pg*, *.fvl*, *.mp4*, *.wmv*, *.mp3*, *.mva*, *.m4a*, *.ogg*, *.av*, *.midi*, por mencionar algunos–.

Cada naturaleza del dato guarda diversos formatos, ya sean binarios, ASCII, de compresión, y pueden procesarse mediante diversas aplicaciones para entregar un producto final. Por ello, es necesario emplear gestores de datos que apoyan en la administración de la información y en el acceso al dato con el objetivo de servir mediante una de interfaz que entregue la información al especialista (investigador) tal como la requiere. Por lo que es preciso considerar protocolos y estándares que intercambien los datos entre aplicaciones de software desarrolladas en lenguajes de programación diferentes, y puedan ser ejecutadas sobre cualquier plataforma, ofreciendo servicios para intercambiar datos que permitan interconectar las estructuras de los datos.

Por lo que es necesario una interconexión para mantener una sincronización de la información, accesible para el usuario mediante los llamados *Webservices* que manejan servicios como:

- Interoperabilidad entre aplicaciones de software, independientemente de sus propiedades o de las plataformas sobre las que se instalen.
- Fomentan los estándares y protocolos basados en texto, que hacen más fácil acceder a su contenido y entender su funcionamiento.
- Apoyan los servicios Web (HTTP).
- Permiten que los servicios y el software de diferentes compañías se combinen fácilmente para proveer servicios necesarios para el tratamiento de los datos.

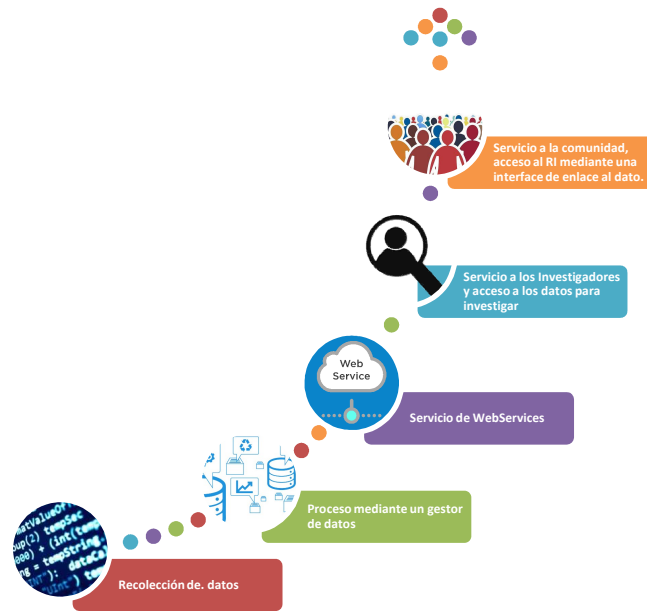


Figura 38. Estructura de software con el enlace del dato.

3.8 Despliegue de la información

El despliegue de la información en el Repositorio debe ser claro, funcional e intuitivo y de fácil extracción. Como se comentó en el capítulo anterior, debe enfatizarse en torno a lo que se quiere dar a conocer, a qué nivel, qué tipo de consultas se necesitan y, sobre todo, la presentación de la información útil al usuario final.

Este despliegue puede basarse en la catalogación del dato o recurso, en la estructura de la comunidad en donde se integró y en el usuario final a quien va dirigido. En la Figura 39 se presenta el orden que se requiere para una estructura de despliegue:



Figura 39. Estructura de despliegue de la información.

Dentro de los Repositorios, establecer un despliegue de información de los elementos permite identificar el registro y el recurso ligado. Estos son los datos que se recomiendan presentar en el primer registro de los datos:

- Autoría.
- Títulos del recurso.
- Elementos geográficos y coordenadas.
- Fechas de creación.
- Ingreso y almacenamiento.
- Temas y/o palabras clave.

Estos elementos, dependerán de la naturaleza del dato de cada institución además el tipo de formato de despliegue y el resto de la información del registro, vendrá en una segunda pantalla o en el registro completo. Es necesario dejar claro los elementos por lo cuales accedan y consulten en una primera vista los usuarios. Presentamos algunos ejemplos de RI de Datos, en donde el despliegue asegura la validez, calidad, seguridad y operatividad, además de que se requieren de variados software para el despliegue, la visualización y la extracción del dato, a fin de que puedan ser reutilizados, si es el caso.

Por favor, use este identificador para citar o enlazar este ítem: <http://rigeofisica.ssn.unam.mx/jspui/handle/ssnmx/67>

Título : Datos de velocidad en continuo a 100 muestras por segundo

Autor : Servicio Sismológico Nacional (IGEF-UNAM)

Palabras clave : CIENCIAS FÍSICO MATEMÁTICAS Y CIENCIAS DE LA TIERRA
CIENCIAS DE LA TIERRA Y DEL ESPACIO
GEOFÍSICA
SISMOLOGÍA Y PROSPECCIÓN SISMICA

Fecha de publicación : 2009

Editorial : UNAM, IGEF, SSN, Grupo de Trabajo

URI : <http://rigeofisica.ssn.unam.mx/jspui/handle/ssnmx/67>

Aparece en las colecciones: Velocidad Continua

Figura 40. Presentación del registro del dato en el RI, en una primera fase en el software de DSpace.⁸²

Figura 41. Estructura de despliegue de la información con un tipo de software para la presentación de los datos en la segunda fase.

⁸² UNAM. SSN. Las figuras 40-41 fueron tomadas del Repositorio Institucional de Datos del Servicio Sismológico Nacional. *Op. Cit.*

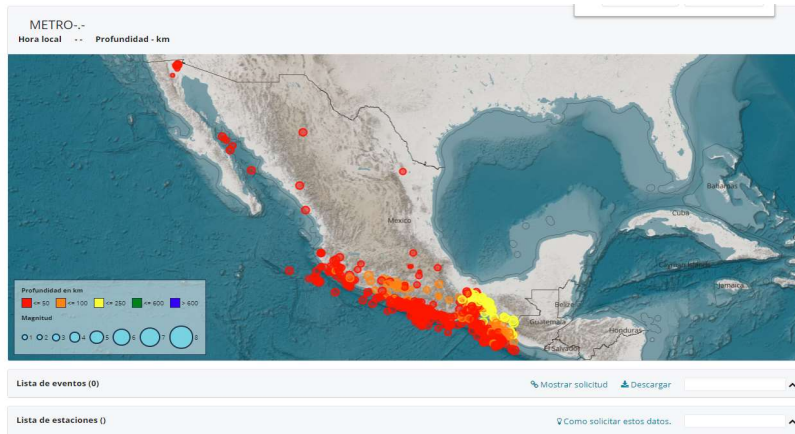


Figura 42. Presentación del software para la consulta gráfica del dato en el RI del SSN.⁸³

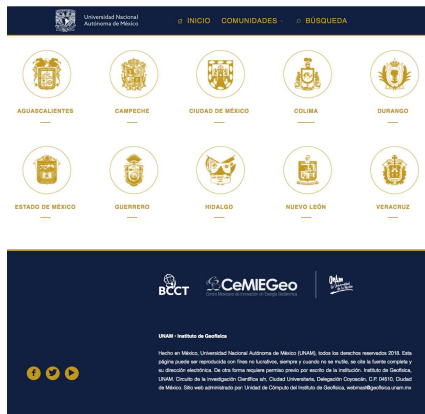


Figura 43. Presentación de la organización de los datos en el RI de Geotermia, en el software de DSpace.

- Artículos
- Carta Hidrogeológica
- Dataset
- Informes Técnicos
- Libros
- Resúmen de Congreso
- Tesis

Figura 44. Estructura de despliegue de la información en el RI de Geotermia.

```

Predeterminado
1 No. Registro, Estado, Nom, Nom_or, Fecha_muestreo, T, pH, CE_mScm, STD_ppm, Na_ppm, K_
2 12, Oaxaca, POZO TLAXIACO, OCBAL2857, 29/04/2015, 19.3, 7.5, 675, 507, <0.3427, 2.147, 1
3 13, Oaxaca, POZO N°6 TAMAZULAPAN, OCBAL2858, 29/04/2015, 21.7, 2.670, 538, 22.31, 2.26
4 14, Oaxaca, POZO HUAJUAPAN, OCBAL2866, 29/09/2015, 20.3, 7.4, 1119, 664, 55.55, 14.51, 1
5 15, Oaxaca, POZO SAN AGUSTIN ATENANGO, OCBAL2868, 23/06/2015, 22.1, 8.3, 389, 278, <0.
6 16, Oaxaca, POZO MARISCALA DE JUAREZ, OCBAL2875, 22/06/2015, 26.3, 7.4, 725, 449, 36.1
7 17, Oaxaca, POZO CALIHUALA (BALSAS), OCBAL2877, 17/08/2015, 22.3, 7.3, 568, 492, , 62.
8 18, Oaxaca, POZO HUITZO, OCPSU4760, 25/08/2015, 22.4, 7.4, 559, 379, 18.62, <0.1124, 75,
9 19, Oaxaca, 202S-T8 SAP UNION ZAPATA, OCPSU4710, 16/10/2015, 24.7, 8.1467, 984, 238.5
10 20, Oaxaca, POZO LA LUZ 5. CODIPSA 37, OCPSU4859, 31/07/2015, 31.4, 6.3, 464, 306, 28.1
11 21, Oaxaca, POZO No. 47 Codipsa 28 MAGDALENA ZAHUATLAN, OCPSU4865, 07/05/2015, 24.
12 22, Oaxaca, POZO EL SAUZ, OCPSU4868, 02/09/2015, 23.8, 7.417, 301, 10.63, 3.979, 42.47,
13 ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,
14 ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,
15 ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,
16 ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,
17 Comisión Nacional del Agua (CONAGUA) | Gobierno | gob.mx, ,,,,,,,,,,,,,,,,,,,,,
18 CONAGUA gerencia de Calidad del Agua datos geoquimicos del Estado de Oaxaca, ,
19 Metodología: Proceso metodologico estructurado para la recopilacion de datos ,

```

Figura 45. Presentación del dato, preparado para su reutilización en formato csv.⁸⁴

⁸³ UNAM. SSN. La figura 42 fue tomada del software que se enlaza con el Repositorio Institucional de Datos del Servicio Sismológico Nacional. *Op. Cit.*

⁸⁴ UNAM. IGEF. Las figuras 43-45 fueron tomadas del RI de Recursos Geotérmicos para México. *Op. Cit.*

1 to 10 of 308 Results

International Late Yellow Hybrid Trial - ILYH0704
 Jul 4, 2019
 Global Maize Program 2019, "International Late Yellow Hybrid Trial - ILYH0704", hdl:11 Software Repository Network, V1
 Summary results and individual trial results from the International Late Yellow Hybrid - ILYH (Elite Hybrid Trial - CHTTY) conducted in 2007.

International Early Yellow Hybrid Trial - IEYH0612
 Jul 4, 2019
 Global Maize Program 2019, "International Early Yellow Hybrid Trial - IEYH0612", hdl:11 Software Repository Network, V1
 Summary results and individual trial results from the International Early Yellow Hybrid - IEYH (Topo conducted in 2006.

Livelihood Impacts of Commercial Maize Production in the Plateau Eco
 Jul 2, 2019
 Keil, Alwin; Mitra, Archisman; Parida, Nabakishore; Ifikar, Wasim. 2019. "Livelihood Impacts of Commercial Maize Production in the Plateau Ecology of Odisha, India", hdl:11529/10548161, CIMMYT Research Data UNF6K6Campr/baBXW/P0iOreHjg--
 Large tracts of land in the upland plateau ecology of Odisha are not suitable for rice production as district of the plateau area, nearly half of the total cultivable area (43%) falls into the upland category.

Filters:
 Repositories (1)
 Datasets & Software (307)
 Files (3195)
Publication Date: 2017 (97), 2016 (90), 2018 (63), 2015 (35), 2019 (20)
Subject: Agricultural Sciences (288), Social Sciences (2), Earth and Environmental Sciences (1)
Series Name: International Maize Trial Network (66)
Keyword Term: Agricultural research (186), Maize (112), Grain yield (75), Triticum aestivum (75), Crop performance (73)
Geographic Coverage Country / Nation: Mexico (100), India (52), Ethiopia (47), Nepal (41), Pakistan (40)

Figura 46. Otro ejemplo de RI de datos, presentación del dato.⁸⁵

Archivos Metadatos Condiciones Versiones

Search this dataset... Encontrar

22 archivos

CHTTY 2007 El procedimiento de árbol Análisis de agrupamiento de varianza mínima de Ward	Descargar
07CHTTY-rendimiento de grano.xls MS Excel - 37.5 KB - 4 de julio de 2019 - 1 Descargar MDS: 337d534edcc2d851bd4af8367969ed79	Descargar
07CHTTY-Locations.xls MS Excel - 30.0 KB - 4 de julio de 2019 - 0 Descargas MDS: 5bc078dcb949ae52ba27a538a39828	Descargar

Figura 47. Estructura de despliegue de la información.

7	Num.	Pedigree	Entry	Yield kg/ha	Yield gnrwtg kg/ha
8	1	CLQ-RCYQ49 = (CML176 x CL-G2501)-B-55-2-1-B*4-B/CML165 +P66C1HC144-3-1-1-BB-4-B*8-B	1	5460.7	4930.8
9	2	CLQ-S89YQ06-S89TYQF9-1-B-6-#-2-1-BB-3-B-1-B*6/CML161+G25QC18MH520-1-1#-1-2#-5-3-B-1-BBBB-#-B*14	2	5866.6	7323.7
10	3	CLQ-RCYQ46=(CML150 x CL-Q3618)-B-17-2-2-BBB-B/CML161+G25QC18MH520-1-1#-1-2#-5-3-B-1-BBBB-#-B*14	3	5522.5	5630.8
11	4	CLQ-S89YQ04+S89TYQF9-1-B-4-#-2-B*8-B/CML161+G25QC18MH520-1-1#-1-2#-5-3-B-1-BBBB-#-B*14	4	5980.3	6388.2
12	5	CLQ-RCYQ49 = (CML176 x CL-G2501)-B-55-2-1-B*4-B/CML161+G25QC18MH520-1-1#-1-2#-5-3-B-1-BBBB-#-B*14	5	5818	6020.1
13	6	CLQ-RCYQ47 = (CML159 x CL-G2501)-B-1-1-2-2-B*4-B/CML161+G25QC18MH520-1-1#-1-2#-5-3-B-1-BBBB-#-B*14	6	5833	5775.5
14	7	CLQ-RCYQ42 = (CML150 x CL-G2501)-B-33-3-1-BBBB-10-B/CML161+G25QC18MH520-1-1#-1-2#-5-3-B-1-BBBB-#-B*14	7	5398.6	5441.6

Figura 48. Presentación del dato.

Es importante que en el registro del dato se permita identificar y acceder al recurso. Consideremos que la estructura del Repositorio de Datos es compleja se basa en la especialización de los datos que son incluidos tales como: taxonomías, familias, tipologías, etc. y se convertirá en un punto de enlace para ofrecer la

⁸⁵ CIMMYT (2019). Las figuras 46-48 fueron tomadas del Repositorio Institucional de Datos y Software de Investigación del Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT), <https://data.cimmyt.org/dataverse/root?q=&types=dataverses&sort=dateSort&order=asc&page=1>. (Consultado: 8-07-2019).

ciencia abierta al mundo académico. En el **Anexo 1** se muestran elementos que fueron considerados en los RI de Datos y Literatura según las necesidades de la institución propietaria.

3.9. Diseño web

En este punto es necesario diseñar la imagen institucional que permita la identificación del Repositorio, así como del contenido que conlleva, todo ello sin perder la esencia de difundir los datos bajo la llamada ciencia abierta.

De esta manera, debemos considerar una estructura que organice y etiquete los entornos de los datos y de la información que se presenta, a fin de facilitar el acceso o lo que conocemos como la localización de la información contenida, y así mejorar su utilidad y aprovechamiento de la misma por parte de los usuarios.

Se comentaron varios puntos clave en el capítulo anterior en torno a que debe considerarse en cuanto al diseño web de un RI de datos, bajo una política de ciencia abierta, por lo que en este capítulo presentamos ya algunos ejemplos con base en la experiencia adquirida en el desarrollo de tres RI en la UNAM.

El sitio web del RI debe diseñarse bajo una estructura de organización lógica de contenidos, poniendo a disposición de los usuarios finales la información correspondiente a su solicitud. En él se considerarán los espacios, la presentación del contenido y el acceso al recurso, así como el despliegue de datos, convirtiendo el sitio web o portal en un puente de comunicación entre el personal que está desarrollando el RI y los usuarios bajo un entorno de información compartida, la cual será aprovechada por los usuarios finales.

Pero qué debemos considerar en el desarrollo del sitio web:

- La combinación entre los sistemas de organización.
- El etiquetado, de la navegación al interior, con la finalidad de ubicar rápidamente los contenidos con el menor esfuerzo por parte del usuario.
- Establecer relaciones o enlaces para facilitar dicha navegación.

En ese sentido se debe trabajar en:

Estructura

Se recomienda utilizar una distribución jerárquica de categorías de información con el objetivo de que el usuario en un solo vistazo encuentre una vía de navegación de todos los servicios que se ofrecen en el RI. Deben evitarse los submenús en lo posible, para facilitar la navegación.

Etiquetado

Se pueden utilizar enlaces contextuales, los cuales se diferencian mediante ligas activas con diferente formato al texto, creación de ligas externas e independientes (que enlacen a los recursos o al objeto que esté relacionado con el registro) u

otras interfaces fuera del sitio para el despliegue de los datos. También es necesario utilizar lenguaje del área especializada, a fin de componer un ambiente familiar en el web.

Navegación

En cuanto al sistema de navegación, se puede hacer de forma profunda y a lo largo del sitio, es decir, debemos procurar una navegación vertical y horizontal, según el nivel o subnivel. Es recomendable que siempre haya una barra de navegación gráfica que permita el acceso a las principales comunidades o colecciones, según sea la jerarquía que se haya empleado. Cada sección en su interior complementa la navegación del etiquetado, además de enlazar los contenidos con otras barras laterales relativas a estas.

En cuanto a las páginas de segundo nivel, se deben incluir accesos a la página inicial, ya sea ubicada dentro de las barras laterales de navegación o en la parte superior. Y en cuanto a la navegación local, se debe permitir articular los contenidos de los diferentes servicios, logrando marcar y hacer identificables los distintos niveles en los que esté construido el Portal.

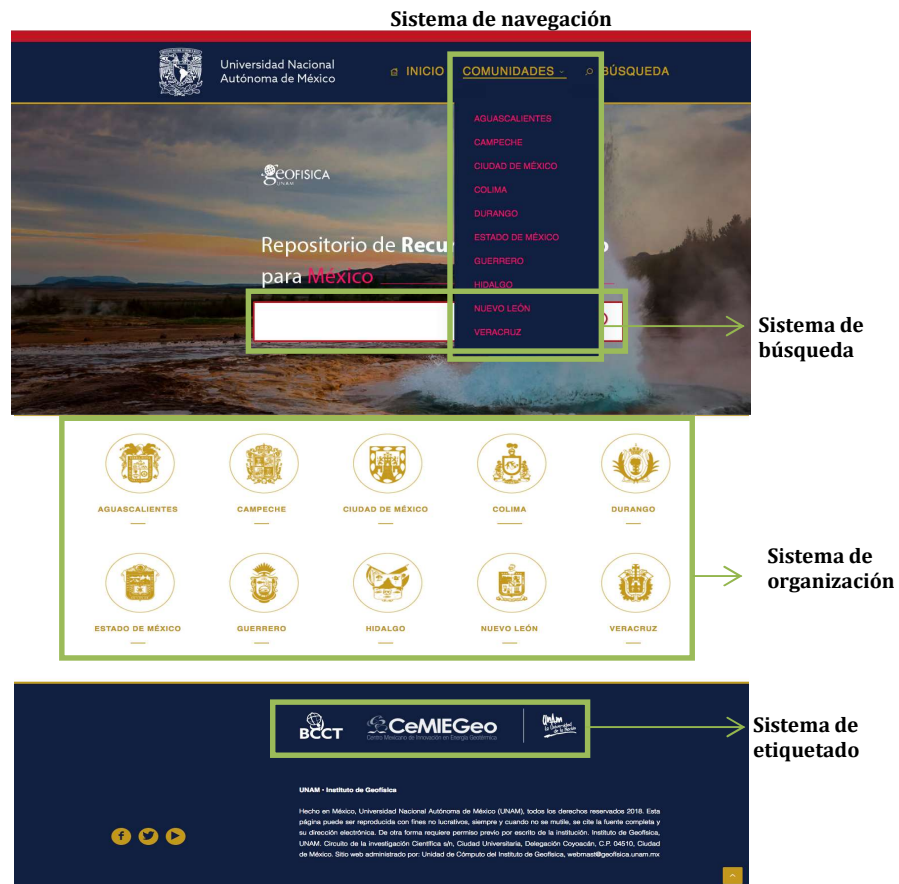


Fig. 49. Diseño web: estructura, etiquetado y navegación.

Accesibilidad, usabilidad y diseño universal

Accesibilidad:

A fin de cubrir la accesibilidad del sitio web, se debe considerar el objetivo central del Repositorio Institucional, el contenido, los recursos, así como las necesidades de información de la comunidad científica que utilizará los datos. Asimismo, deben contemplarse las siguientes características:

- Interface simple e intuitiva, buscando sea familiar para el usuario final.
- Con base en la cantidad de herramientas tecnológicas que existen, se debe ofrecer una correcta visualización en distintos dispositivos desde ordenadores de escritorio a tabletas y móviles redimensionando la colocación de los elementos en la web de forma que se adapten a la medida de cada dispositivo permitiendo una correcta visualización y una mejor experiencia de usuario.. Esto permitirá reducir el tiempo de desarrollo evitando información duplicada, aumentando la viralidad de los contenidos y compartirlos de forma mucho más rápida y natural
- Considerar la compatibilidad con los diferentes navegadores y cambios de plataformas y sistemas operativos.
- El sitio debe poder validarse con estándares internacionales web –como, por ejemplo, W3C–.

Usabilidad:

En cuanto a este punto, se tomó como elemento primario lo que implica la base de la arquitectura de la información en sus diferentes sistemas, además de incluir servicios estructurados, buscando principalmente que los usuarios de nuestra comunidad los reconozcan y aumente el consumo y el uso los contenidos ofertados, garantizando la calidad de información de un modo fácil en cuanto a su localización.

En este punto, no se pueden dejar de lado servicios de estadísticas para monitorear el tráfico, el comportamiento, la información geográfica, uso de navegadores, tipos de dispositivos, etc., independientemente de las estadísticas que se llevaran mediante el Repositorio Nacional, así como el que lleva el propio software del Repositorio.

Posicionamiento Web del Portal

Con base en lo anterior, solo resta brindar una buena localización del portal así como del contenido, para lo cual se pueden aplicar diversas técnicas, logrando que los buscadores indexen cada uno de los sitios y subsitios de los que se compone el portal y puedan posicionarlo en categorías privilegiadas dentro de los primeros resultados en el momento que se realicen las búsquedas. A continuación, presentamos cómo se pueden aplicar algunas técnicas:

- Generar un Sitemaps XML para lograr un buen posicionamiento en los motores de búsqueda estructurado bajo un lenguaje XML. Puede utilizar el siguiente sitio para su generación: <http://www.xml-sitemaps.com/>.

- Implementar Meta Tags. Estas etiquetas que se incluyen en el HEAD: *“Title”, “Keywords”, “Description”, “Language”, “Distribution”, “Author”, “Reply-To”, “Revisit-after”, “Robots”, “Resource-type”, “Content-type”, “Date”, “Rights”*.
- Agregar Google Analytics sería otra opción, pues actualmente la mayoría de las personas cuentan con correo electrónico de la compañía Google y nos permitirá contar con una visión más amplia de su uso.
- Incluir GoSquared, aplicación que monitoreará el tráfico mediante la cantidad de visitantes, país, dirección IP, sistema operativo y navegador, además entrega información diaria de la actividad en el portal.
- Configurar el archivo robots.txt, con base en el tipo de acceso que se requiera en cada uno de los sitios y subsitios de los que se compone el Portal del Repositorio Institucional.

La visibilidad de un sitio web y su posicionamiento son elementos clave que permitirán al Repositorio contar con una mayor consulta por parte de los usuarios. Debemos entender que el Portal del RI tiene que ser un ente activo en el que debemos pensar en estrategias que faciliten su acceso y consulta.

CONCLUSIONES

Trabajar en el desarrollo de un Repositorio Institucional de Datos en donde la aportación de la experiencia adquirida en la práctica –en conjunto con la documentación académica existente– es de gran beneficio en el impacto científico y bibliotecológico, en donde en este último se centró la investigación estableciendo los elementos necesarios que facilitaron el desarrollo de un Repositorio Institucional de Datos para la investigación Científica el cual apoya a la investigación especializada en Ciencias de la Tierra, bajo lineamientos y estándares que permitieron una interoperabilidad e integrarnos en un Proyecto de alcance nacional.

Pues refleja el desarrollo de nuevos proyectos de investigación y elimina en gran medida la duplicidad de esfuerzos y tiempo, logrando con ello aumentar el conocimiento del uso y manejo de la información especializada en las Ciencias de la Tierra, producida en México, si consideramos que existen aproximadamente 27,188⁸⁶ investigadores con base en los datos proporcionados por el SNI, de los cuales 4,245 son especializados en Ciencias de la Tierra, el Mar y la Atmósfera, con niveles de estudios que van desde licenciatura hasta doctorado, de los cuales 1,150 pertenecen a la UNAM; 222, al IPN; de la UAM, 189; 158, al CINVESTAV; 112, al Centro de Investigación Científica y Educación Superior en Ensenada (CICESE); 64, de la Universidad Autónoma de Baja California (UABC); 111 a la Universidad de Guadalajara (UdeG), y el resto a otras instituciones en el país.

Del total de la producción científica indizada el Web of Science (WOS) en Ciencias de la Tierra en México, el 27.09% es generada por la UNAM; el 5.22, por el IPN; el 4.45, por la UAM; 3.72, por el CINVESTAV; el CICESE produce el 2.63; la UdeG, 2.61; y el 1.50% por la UABC, y, como vemos, la Universidad Nacional cuenta con la mayor producción científica.

Cabe mencionar que las Ciencias de la Tierra a nivel nacional cuentan con una importante participación en el fortalecimiento de la investigación y en la colaboración mundial con otras entidades académicas, convirtiéndose a lo largo de los años en una de las principales ciencias emergentes del país, y con una cooperación con casi todos los países del mundo.

El desarrollado en este trabajo concluye en los siguientes aspectos:

- Se mostró la creación de un Repositorio Institucional de Datos que pertenecen al área científica de ciencias de la tierra, según su naturaleza, basado en la Iniciativa de integración de Acceso Abierto de la información científica y tecnológica del Capítulo 10 de Ciencia y Tecnología y en los *Lineamientos Generales para el Repositorio Nacional (RN) y Repositorios Instituciones (RI)*, aumentando la visibilidad y participación en proyectos de

⁸⁶ México. Conacyt (2019).“Tabla de Investigadores Vigentes 2017 por Grado, Nivel, Género, Adscripción, Entidad Federativa y Área de Conocimiento”. Disponible en <https://datos.gob.mx/busca/dataset/sistema-nacional-de-investigadores>. (Consultado: 8-07-2019).

alcance nacional bajo un contexto bibliotecológico y que además apoya a la investigación en Ciencias de la Tierra e impulsa la estrategia nacional para democratizar la información.

- Se logró bajo la política nacional de información científica y de lineamientos de alcance nacional ofrecer más allá de lo que brindan las publicaciones tradicionales en el Repositorio de datos mediante el acceso a datos puros.
- Como se demuestra al inicio de las conclusiones dentro del ámbito de las Ciencias de la Tierra la producción de los datos crece de manera exponencial por ser datos continuos, ininterrumpidos y no repetibles que requieren de ser procesados, es por eso que se describen en el capítulo 3 una serie de estrategias que apoyan la gestión, distribución y difusión de los datos de forma interoperable para ser localizables y se genere nuevo conocimiento en esta especialidad bajo las políticas de Acceso Abierto y como resultado potenciarlas y aumentar la visibilidad a nivel nacional e internacional.
- Otro de los aspectos que se constató que es importante trabajar en este tipo de proyectos para aumentar la conformación de Repositorios Institucionales a mediano plazo en México, para dar a conocer la ciencia de manera abierta, debido a que –con base en la información del Directorio de Repositorios de Acceso Abierto OpenDOAR– solo contamos con 41 Repositorios registrados. En América Central se tienen 66. Por ello, necesitamos contribuir a que la información producida en México esté accesible y sea de consulta de Acceso Abierto haciéndose necesaria la creación de mecanismos que apoyan la difusión de la producción científica.
- Como se expone en el capítulo 2 y 3 la generación este tipo de plataformas es un hecho que cambia el hábito, la consulta de la publicación y el manejo de información entre los usuarios, que repercute indudablemente en nuestra especialidad la bibliotecología, lo que nos está obligado a innovarnos en nuestros perfiles profesionales y trabajar de manera conjunta con otros especialistas en diferentes disciplinas siendo un punto clave para ocuparnos por la ciencia abierta y lograr establecer un producto de impacto nacional.

Por otro parte en el panorama actual de la ciencia mundial, México debe ganar un lugar y lograr una comunicación clara con otros repositorios generados en el mundo, creando un proyecto competitivo. Por lo tanto, la aportación que se hace permite a otros organismos seguir una serie de pasos y experiencias que facilitan la realización de un RI de manera más fácil y con un mejor entendimiento.

En el capítulo dos, se presentan las diferencias entre un *Repositorio de Datos de Investigación (RDR, Research Data Repositories)* el cual pertenece a una institución que cuenta con recursos y servicios utilizados por la comunidad científica para realizar investigaciones que, por lo regular, el tipo de datos que se publican en el RDR se asocia o se encuentran respaldados y vinculados a una publicación o proyecto como resultado final.

Por otro lado se encuentra el *Repositorio de Datos para la Investigación (DRR, Data Repositories for Research)*, esta herramienta contiene datos puros y continuos de apoyo a investigaciones o para investigar nuevas propuestas y tienen la condición de que sus datos pueden ser reutilizados para diferentes fines, intereses o del surgimiento de un nuevo fenómeno y/o propósitos que puedan generar conocimiento y descubrir otras líneas de investigación, además trabajan bajo una dinámica diferente y relacionar a grupos multidisciplinarios de especialistas del área científica. Por mencionar algunos ejemplos, como grabaciones de eventos o análisis y observaciones del estado de la Tierra, del Sistema Solar en tiempo real (datos no repetibles que nacen a cada momento), permitiendo la investigación en distintas disciplinas, como la economía, la protección civil y, sobre todo, las diversas líneas de investigación a las que puede servir, a fin de ir más allá de lo que nos ofrecen actualmente las publicaciones que tradicionales.

En el capítulo tres presento el desarrollo por etapas para planear y llevar a cabo desarrollo de un Repositorio Institucional de Datos, por la naturaleza de la ciencias de la tierra los datos requieren de un tratamiento para su acceso, publicación y reúso. Es por eso que se presenta dos tipos de Repositorios en donde se aportan elementos que pueden apoyar en la orientación de su desarrollo y alojen los datos y se compartan convirtiéndose ya en una práctica habitual en ésta disciplina.

Trabajar en el desarrollo de un Repositorio de manera conjunta con diversos especialistas maximiza la visibilidad, el impacto de la producción científica, académica e institucional y aumenta su crecimiento sin que su contenido se duplique con otras fuentes secundarias de información, como son las bases de datos y los catálogos, logrando proyectar nuevas perspectivas y, al mismo tiempo, dar visibilidad a la información científica, haciéndola disponible para un grupo mayor de usuarios de distintos niveles académicos.

Es importante aclarar y diferenciar y sobretodo entender que un Repositorio no es una base de datos sino la puerta para ofrecer la ciencia abierta al mundo académico mediante un modelos de operación que cumple estándares internacionales con el propósito de enfocarse a tres líneas de acción: el almacenamiento, la preservación y la diseminación de la información académica, científica, tecnológica y de innovación, la cual se deriva de las investigaciones en lo que se refiere a datos. Bajo el cumplimiento de los Principios FAIR con miras a hacerlos localizables e interoperables mediante metadatos, accesibles con el uso de protocolos y, sobre todo, reutilizables bajo licencia abierta, facilitando el intercambio de la información producida en todas las ramas del conocimiento.

Con la finalidad de mostrar un panorama amplio y de contexto nacional el resultado de cómo trabajan las instituciones del país en el desarrollo de Repositorios Institucionales de datos, se llevó a cabo un estudio en donde se revisaron 96 RI de México sobre una serie de elementos que consideraron cada una de las instituciones que forman parte del Proyecto Nacional *Programa de*

Repositorios del Consejo Nacional de Ciencia y Tecnología (Conacyt). De aquí, se extrajeron las características más preponderantes de los RI de Literatura y Datos, desde un punto de vista bibliotecológico. El resultado que se obtuvo se encuentra descrito a detalle en el <https://cutt.ly/myvKLDp> de este trabajo, así como la graficación de la información en el siguiente orden:

1. Se llevó a cabo el análisis y la organización de la información. Se detalla el tipo de formato y conformación que alberga cada RI (**ver anexo 1**).
2. Se presenta el Tipo de Repositorios Institucionales (**ver anexo 3**).
3. Se revisaron los permisos y licenciamientos en cada plataforma RI, así como el tipo de servicios REST que utilizan (**ver anexo 4**).
4. Se reconocieron los diversos tipos de identificadores que se utilizan en cada uno de los RI (**anexo 5**).
5. Se examinó el tipo de metadatos de cada Repositorio Institucional (**ver anexo 6**) así como su visualización y accesibilidad (**ver anexo 7**).
6. Dentro de esta auscultación se tomaron en cuenta las áreas temáticas que utilizaron cada tipo de RI, independientemente de las que solicita Conacyt (**ver anexo 8**).
7. Se revisó la estructura de los dominios, la personalización del acceso y el identificadores llamado Handle de cada RI (**ver anexos 9 y 10**).
8. Finalmente, también revisamos el diseño, la estructura de la búsqueda y el despliegue de los datos, así como la imagen institucional y el acceso desde el RN (**ver anexos 11 y 12**).

Concluyo que el contar con un Repositorio Institucional de datos fomenta el desarrollo de la ciencia y la investigación. En nuestro proyecto, de manera particular, pone al día tanto a especialistas como al público en general con información diseminada y expedita, la cual orienta la toma de decisiones en cuestiones ambientales y climáticas por mencionar algunos elementos prioritarios en México, por su amplio territorio y variedad de áreas geográficas, brindando apoyo a la protección civil.

Bajo el panorama anterior, debemos ocuparnos más en que los RI pasen a ser considerados como un beneficio en el desarrollo de la ciencia. En la actualidad no todo se encuentra en Acceso Abierto y aún es poco claro el intercambio de datos generados por organizaciones académicas, entidades y diversos organismos, pero se debe enfatizar más en trabajar en ellos en colaboración con otras instituciones, desarrollando proyectos de amplio espectro y, sobre todo, con información útil para los usuarios.

Lo que es un hecho: la ciencia se está moviendo hacia una mayor apertura en términos de un aumento de datos, publicaciones, software y en flujos de trabajo. Sin embargo, los investigadores están aprendiendo a navegar y a los cambios que surgen en la forma de publicar sus investigaciones en el campo de la ciencia abierta, enfrentándose, de igual forma que los bibliotecólogos, a nuevas formas de trabajo y al juicio de dejar abierta la información para una consulta masiva.

Pensar también en aquellos que desean que sus datos sean más abiertos muestra la variedad de opciones sobre dónde y cómo compartirla. Es necesario ahondar en el conocimiento, en la curación de datos y metadatos (información que describe un conjunto de datos). Esto ayuda a garantizar que los datos sean útiles para quienes quieren compartir. Por lo tanto, al desarrollar un RI debe pensar a corto y mediano plazo, bajo una estructura que permita ser un sitio en donde los investigadores confíen, publiquen y se pueda almacenar y disponer de sus trabajos de investigación, al mismo tiempo que las instituciones los consideren como verdaderas herramientas de apoyo a la investigación y a la sociedad en general.

Esta apertura, en cuanto a datos se refiere, produce beneficios como la creación de nuevas colaboraciones con diversos grupos de trabajo y de especialidades, aumentando con ello la confianza en los hallazgos, lo que genera buena voluntad entre los investigadores para ingresar los resultados de sus investigaciones.

Si bien no tenemos claro que tan rápido avanzará este proceso, por ellos debemos seguir trabajando de forma conjunta y colaborativa para doblar esfuerzos y mostrar los beneficios y alcances que estas plataformas ofrecen bajo la política de la ciencia abierta.

Finalmente, es importante señalar y recalcar que es imprescindible la participación de profesionales de la información que apoyen con sus conocimientos académicos y experiencias, que se encuentren inmersos en este proceso. Es una buena oportunidad de posicionarnos y adentrarnos a otros campos de la ciencia, impulsando nuestra profesión, además de ser punto clave para trabajar de forma colaborativa y enlazar a las personas para ofrecer una red de comunicación, generando nuevos proyectos de ciencia abierta. Debemos estar atentos a los cambios que se presentan en los datos y a las múltiples posibilidades de su uso, debido al gran impulso que están tomando para la investigación a nivel mundial, sin descuidar el principal propósito. Hoy en día, los datos para la investigación es un tema imprescindible para los bibliotecarios y demás especialistas de la información.

BIBLIOGRAFÍA

- Alonso, J. S.; Imma y Ma. Luisa Martínez Conde (2008). *Informe APEI sobre Acceso Abierto: APEI*.
- American Geophysical Union (2019a). "Declaration adopted by the American Geophysical Union". Disponible en <https://ethics.agu.org/>.
- American Geophysical Union (2019b). "Repository Finder". Disponible en <https://repositoryfinder.datacite.org/>.
- Arlitsch, K. (2018). "Why So Many Repositories? Examining the Limitations and Possibilities of the Institutional Repositories Landscape". *Journal of Library Administration*, 58(3), 264-281.
- Arlitsch, K. y. P. O. (2013). *Improving the Visibility and Use of Digital Repositories through SEO: ALA Neal-Schuman LITA*.
- arXiv.org. (2019). Disponible en <https://arxiv.org/>.
- Barrueco Cruz, J. M.; Aurelia Rodríguez; Pilar Rico Castro; Ma. Ángeles Coslado Bernabé (2017). *Guía para la evaluación de repositorios institucionales de investigación*.
- Béjar, E. (2012). Datos abiertos. Disponible en <https://5stardata.info/es/>.
- Bluh, P. et al., (2013). *The institutional repository: benefits and challenges: ALCTS*.
- BOAI Forum (2018). *Budapest Open Access Initiative*. Disponible en <https://www.budapestopenaccessinitiative.org/translations/spanish-translation>.
- Borgman, C. L.; Jillian C. Wallis; Matthew S. Mayernik (2012). "Who's Got the Data? Interdependencies in Science and Technology Collaborations". *Computer Supported Cooperative Work (CSCW)*, V. 21(6), 485–523.
- Callicott, B.; David Scherer; A. Wesolek (2016). *Making Institutional Repositories Work*: Purdue University.
- Conacyt (2017). *Lineamientos específicos para Repositorios*: Conacyt.
- Conacyt (2017). *Lineamientos generales de Ciencia Abierta*: Conacyt.
- Conacyt (2017). *Lineamientos jurídicos de ciencias Abierta*: Conacyt.
- Conacyt (2019). "El Repositorio Nacional de México". Disponible en

<http://www.repositorionacionalcti.mx/>.

Crow, R. (2002). "The Case for Institutional Repositories: A SPARC Position Paper". *Scholarly Publishing and Academic Resources Coalition*, 223, 1-37.

Dallmeier-Tiessen, S. *et al.*, (2010). The Study of Open Access Publishing Consortium, SOAP.

DCD (2019). "Digital Curation Centre". <http://www.dcc.ac.uk/>.

Díaz Pérez, M.; D. Armas Peña; Rodríguez Font; A. F. Camacho Díaz (S/F). "Política informacional para Repositorios institucionales en universidades". *Atlante. Cuadernos de Educación y Desarrollo*.

Doria, M. V.; Ana Maria del Prado; Ma. Carolina Haustein (2014). *Repositorios digitales de Acceso Abierto: Presencia de repositorios en universidades nacionales argentinas*: Editorial Académica Española.

Drake, M. A. (2004). "Institutional Repositories: Hidden Treasures". *Information Today, Inc.*, Vol. 12 (5).

EIFL. (2019). *Electronic Information for Libraries*. Disponible en <https://www.eifl.net/>.

Enciclopedia Libre Wikipedia.

EPSRC (2019). *Engineering and Physical Sciences Research Council*. Disponible en <https://epsrc.ukri.org/about/standards/researchdata/scope/>.

Ferreras Fernández, M. (2016). *Visibilidad e impacto de la literatura gris científica en repositorios institucionales de Acceso Abierto. Estudio de caso bibliométrico del repositorio*. (Tesis Doctoral). Gredos de la Universidad de Salamanca.

García-Moreno, T. H. P.; M. A. (2013). "Datos abiertos y repositorios de datos: nuevo reto para los bibliotecarios". *El profesional de la información*, V. 22 (3).

Ginsparg, P. (2011). "ArXiv at 20". *Nature*, Vol. 476, pp. 145-147.

Global Open Access Portal. (2019). Global Open Access Portal. Disponible en <http://www.unesco.org/new/en/communication-and-information/portals-and-platforms/goap/>.

Gonzales, B. M. (2018). *CLIPP 44: Institutional Repositories*.

González Pérez, L. (2017). "Identidad digital 2.0: Posibilidades de la gestión y

visibilidad científica a través de repositorios institucionales de Acceso Abierto”. En *Congreso Internacional de ecosistemas del conocimiento abierto*.

González Videgaray, M. (2016). *Censo de Herramientas de Gestión de Cursos y Repositorios Digitales para el Aprendizaje en la UNAM*. Disponible en <http://www.plataformas.unam.mx/>.

Gutiérrez Gallego, M. (2015). *Data Sharing: Análisis de repositorios de datos de investigación*. (Trabajo Fin de Grado en Información y Documentación, curso 2014-2015). Disponible en <http://hdl.handle.net/10366/129664>.

Hanley, T. (2015). *Research knowledge at the heart of development? Trends, challenges and opportunities*. Disponible en INASP.

Hodson, S. (2018). “Making FAIR Data a Reality... and the Challenges of Interoperability and Reusability”. En *10th Annual European DDI User Conference, Annual European DDI User Conference*.

Hub, T. K. (2019). “The DataCite Metadata Schema. What is the DataCite metadata schema?”. Disponible en <https://project-thor.readme.io/docs/datacite-metadata-schema>.

ICML (2005). *Declaración de Salvador sobre Acceso Abierto: la perspectiva del mundo en desarrollo*. Disponible en <http://www.icml9.org/public/documents/pdf/es/Dcl-Salvador-AccesoAbierto-es.pdf>

IFLA (2018). *IFLA Statement on Open Access to Scholarly Literature and Research Documentation*. Disponible en <https://www.ifla.org/publications/ifla-statement-on-open-access-to-scholarly-literature-and-research-documentation>.

INFOTEC (2018). *Manual de interoperabilidad con el metabuscador del Repositorio Nacional vers. 3*.

Kaye, J. (2015). “Storing and sharing research data after the ‘Space Race’”. Disponible en <https://www.jisc.ac.uk/blog/storing-and-sharing-research-data-after-the-space-race-25-feb-2015>.

Kitchin, R. (2015). *Funding models for Open Access Repositories*. Maynooth University Research Archive Library. DOI: 10.3318/DRI.2015.4.

Lorenzo Gil, E. (2015). *Estudio de la integración de repositorios en el sistema científico-investigador: alternativas y estado actual: OS-Repositorios*.

Lynch., C. A. (2003). “Institutional Repositories: Essential Infrastructure for

Scholarship in the Digital Age". *Portal Libraries and the Academy*, Vol. 3 (Núm. 2). DOI: 10.1353/pla.2003.0039.

Madrazo, L. (2010). "El conocimiento arquitectónico en la era de la información: los repositorios digitales". *Revista de Edificación*, Vol. 39-40.

Mentzas G. A., Dimitris; Ronald Young y Andreas Abecker (2001). "Knowledge networking: a holistic solution for leveraging corporate knowledge". *Journal of Knowledge Management*.

National Science Foundation (2019). *National Science Foundation Grant General Conditions (GC-1)*. Disponible en <https://www.nsf.gov/pubs/2001/gc101/gc101rev1.pdf>

Nieminen, S. M. (2008). "Bradford Scholars: An institutional repository for the University of Bradford". *ALISS Quarterly*, 3(4).

OpenMINTED (2019). "Open Mining Infraestructure for text & Data (OpenMINTED): Guedelines". Disponible en https://guidelines.openminted.eu/guidelines_for_providers_of_sw_resources/recommended_schema_for_sw_resources.html

Pampel, H. *et al.* (2013). "Making Research Data Repositories Visible: The re3data.org Registry". *PlosOne*. DOI: 10.1371/journal.pone.0078080.

Real Academia Española (2001). *Diccionario de la Lengua Española*. Madrid, (22a. ed.): Real Academia Española.

Rodrigues, E. (2017). "Next Generation Repositories: Behaviours and Technical Recommendations of the COAR". Disponible en <https://www.coar-repositories.org/activities/advocacy-leadership/working-group-next-generation-repositories/>.

S/A (2004). *Washington D.C. Principles For Free Access to Science A Statement from Not-for-Profit Publishers*. Disponible en <http://www.dcprinciples.org/statement.pdf>.

S/A (2004). *Valparaiso Declaration for Improved Scientific Communication. Strengthening Editors and Scientists Capabilities in Electronic Publishing*.

S/A (2014). "The Subversive Proposal at 20". Disponible en <https://poynder.blogspot.com/2014/06/the-subversive-proposal-at-20.html?sref=tw>.

S/A (2018). *Declaración de Panamá de Ciencia Abierta (2018)*. Diposible en https://hiperderecho.org/wp-content/uploads/2018/11/declaracion_panama_ciencia_abierta.pdf

- S/A (2019). "Data Converter. Convierte datos en varios formatos compatibles con la web, incluidos HTML, JSON y XML". Disponible en <https://shancarter.github.io/mr-data-converter/>.
- Sale, A. (2006). "Comparison of content policies for institutional repositories in Australia". *First Monday*, Vol. 11(4-3).
- Segob (2019). *Datos Abiertos 100 México*. Disponible en <http://www.opendata500.com/mx/>
- Segob (2019). *Datos abiertos de México*. Disponible en <https://datos.gob.mx/>.
- Serrano Vicente, R. (2017). *Evaluación de los repositorios institucionales de Acceso Abierto en España*. (Tesis Doctoral), Universitat de Barcelona.
- Source, F. G. O. (2019). "GeoNetwork opensource". Disponible en <https://geonetwork-opensource.org/manuals/3.6.x/es/index.html>
- Suber P. B., Patrick *et al.* (2003). "Bethesda Statements on Open Access Publishing". Disponible en https://www.researchgate.net/publication/48547523_The_Bethesda_Statement_on_Open-Access_Publishing.
- Tavares, R. (2017). *Implications of Open Access Repositories Quality Criteria and Features for Teachers' TPACK Development*: Springer.
- Torres-Salinas D., Nicolás; Robinson García; Álvaro Cabezas-Clavijo (2012). "Compartir los datos de investigación en ciencia: introducción al datasharing". *El profesional de la información*, 2(12).
- UNAM (2019). *Portal de Datos Abiertos UNAM: colecciones universitarias*. Disponible en <https://datosabiertos.unam.mx/>
- UNAM. IGEF (2019). Repositorio Institucional de Recursos Geotérmicos para México. Disponible en <http://rigeotermia.geofisica.unam.mx/jspui/>
- UNAM. SSN (2019). Repositorio Institucional de Datos del Servicio Sismológico Nacional. Disponible en <http://rigeofisica.ssn.unam.mx/jspui/>
- Vickers, A. J. (2006). "Whose data set is it anyway? Sharing raw data from randomized trials". *Trials*. DOI: 10.1186/1745-6215-7-15.
- Vinit, K. (2018). "A Model for Content Enrichment of Institutional Repositories Using Linked Data". *Journal of Web Librarianship*, 12(1), 46-62. DOI: 10.1080/19322909.2017.1392271.

Vrana, R. (2011). "Digital repositories and the future of preservation and use of scientific knowledge". *Informatologia*, 44(1), 55-62.

Webster, N. (2019). *Merriam-Webster's Learner's Dictionary*.

Wellcome., S. H. (2019). "Open Access Policy". Disponible en http://www.wellcome.ac.uk/doc_WTD002766.html/.

Anexo 1. Cuadro comparativo de los Repositorios Institucionales que interoperan con el Repositorio Nacional hasta junio de 2019

En este Anexo, se puede observar el resultado de la revisión de 96 Repositorios Institucionales de Datos y Literatura, de estos resultados, los de Literatura representan el 87% de todos los que interoperan. Se muestran además características sobresalientes como el análisis y la organización así como los elementos que se extrajeron de cada uno de ellos.

El tamaño del documento es Tabloide con el fin de que se pueda apreciar toda información. Se encuentra en formato del archivo PDF.

Disponible en: <https://cutt.ly/myvKLDp>

ANEXO Núm. 1
Cuadro comparativo de los Repositorios Institucionales que interoperan con el Repositorio Nacional hasta el 30 de junio de 2019

NUM.	INSTITUCIÓN	NOMBRE DEL RI	ACRÓNIMO	TIPO DE RI LITERATURA - DATOS	SOFTWARE	COMUNIDAD DIRIGIDA	TIPOLOGÍA DE OBJETOS DIGITALES	IDENTIFICADOR DEL RECURSO	IDENTIFICADOR PERSONAL	TEMAS CONACYT	ÁREAS TEMÁTICAS PROPIA	LICENCIAS	ORGANIZACIÓN COMUNIDADES	ACCESO AL OAI	URL OFICIAL	BUSQUEDA	TIPO DE INTERFAZ
1	Infotec Centro Público de Investigación INFOTEC	Repositorio Institucional INFOTEC (RII)	INFOTEC	Lit	Dspace	Alumnos, Profesores, Investigadores, Científicos, Tecnólogos y la Población en General a Través de la Difusión del Conocimiento.	Libros, Artículos, los Proyectos de Titulación, Patentes, Memorias de Eventos e Información Referente a los desarrollos Tecnológicos que se realizan dentro de la Institución.	/handle/1027	ORCID	Temas CONACYT	LEMB	BY-NC-ND	Formato	Si	https://infotec.repositorioinstitucional.mx/jspui/	Busqueda Simple y Avanzada	gob.mx
2	Centro de Investigación y Docencia Económicas, A.C.	Repositorio Digital CIDE	RDC	Lit	Dspace	Alumnos, Profesores, Investigadores	informes, Grabaciones de Audio, Documentos de Trabajo, Programs, Tesis, Cuadernos, Revistas	Handle/11651	CURP	Temas CONACYT	No	CC BY-NC-ND	Formato	No	https://cide.repositorioinstitucional.mx/jspui/	Busqueda Simple y Avanzada	gob.mx
3	Consejo Nacional de Ciencia y Tecnología	Repositorio Nacional	rndirecto												https://www.repositorionacionalcti.mx/		
4	Centro de Investigaciones y Estudios Superiores en Antropología Social	Repositorio Institucional del CIESAS	REPOSITORIOCIASAS	Lit	Dspace	Students	Artículos de Revistas y Tesis	Handle/1015	CURP	Temas CONACYT	URN	BY-NC-SA	Formato	Si	https://ciesas.repositorioinstitucional.mx/jspui/	Busqueda Simple y Avanzada	gob.mx
5	Centro de Investigación Aplicada en Tecnologías Competitivas	Repositorio Institucional CIATEC	RICIATEC	Lit	Dspace	Alumnos, Profesores, Investigadores	Artículos, Capítulos de Libros, Patentes y Tesis	Handle/1019	CURP	Temas CONACYT	LEMB	BY-NC	Formato	Si	https://ciatec.repositorioinstitucional.mx/jspui/	Busqueda Simple y Avanzada	gob.mx
6	Instituto Potosino de Investigación Científica y Tecnológica A.C.	Repositorio IPICYT	IPICYT	Lit	Dspace	Alumnos, Profesores, Investigadores	Informes, Documentos, Publicaciones Científicas	Handle/1027	CURP	Temas CONACYT	No	BY-NC-ND	Área	Si	https://ipicyt.repositorioinstitucional.mx/jspui/	Busqueda Simple y Avanzada	gob.mx
7	El Colegio de San Luis, A.C.	Repositorio COLSAN	RCOLSAN	Lit	Dspace	Alumnos, Profesores, Investigadores	Tesis, Artículos, Resúmenes y Libros	Handle/1013	CURP	Temas CONACYT	LEMB	BY-NC-ND	Área / Formato	Si	https://colsan.repositorioinstitucional.mx/jspui/	Busqueda Simple y Avanzada	gob.mx
8	Instituto de Ecología, A.C.	Repositorio Institucional INECOL	RINECOL	Lit	Dspace	Alumnos, Profesores, Investigadores	Artículos, Tesis	Handle/1005	CURP	Temas CONACYT	No	BY-NC-ND	Área / Formato	Si	https://inecol.repositorioinstitucional.mx/jspui/	Busqueda Simple y Avanzada	gob.mx
9	Centro de Investigación Científica de Yucatán, A.C.	Repositorio Institucional CIICY	CIICY	Lit	Dspace	Alumnos, Profesores, Investigadores	Artículos, Informes Técnicos, Patentes y Tesis	Handle/1003	ORCID	Temas CONACYT	No	BY-NC-ND	Área / Formato	Si	https://ciicy.repositorioinstitucional.mx/jspui/	Busqueda Simple y Avanzada	gob.mx
10	Instituto de Investigaciones Dr. José María Luis Mora	Repositorio Mora	MORA	Lit	Dspace	Alumnos, Profesores, Investigadores	Tesis y Revistas, Capítulos de Libros, Documentos	Handle/1018	CURP	Temas CONACYT	LEMB	BY-NC	Área	Si	https://mora.repositorioinstitucional.mx/jspui/	Busqueda Simple y Avanzada	gob.mx
11	Centro de Investigación en Química Aplicada	Repositorio Institucional del Centro de Investigación en Química Aplicada	RIQIQA	Lit	Dspace	Alumnos, Profesores, Investigadores	Capítulo de Libro, Casos de Estudio, Informes Técnicos y Libros y Tesis	Handle/1025	ORCID	Temas CONACYT	No	BY-NC-ND	Formato	Si	https://ciqa.repositorioinstitucional.mx/jspui/	Busqueda Simple y Avanzada	gob.mx
12	Centro de Investigación y Asistencia en Tecnología y Diseño del Estado de Jalisco, A.C.	Repositorio Institucional de CIATEJ	CIATEJ	Lit	Dspace	Alumnos, Profesores, Investigadores	Propiedad Intelectual, Publicaciones y Tesis	Handle/1023	CURP	Temas CONACYT	No	BY-NC-ND	Formato	Si	https://ciatej.repositorioinstitucional.mx/jspui/	Busqueda Simple y Avanzada	gob.mx
13	Instituto Nacional de Astrofísica, Óptica y Electrónica	Repositorio Institucional del INAOE	inaoeRI	Lit	Dspace	Alumnos, Profesores, Investigadores	Artículos, Reportes Técnicos, TesisDoY y TesisMa	Handle/1009	CURP	Temas CONACYT	No	BY-NC-ND	Formato	Si	https://inaoe.repositorioinstitucional.mx/jspui/	Busqueda Simple y Avanzada	gob.mx
14	Facultad Latinoamericana de Ciencias Sociales Sede México	Repositorio Institucional FLACSO México	flacsorep	Lit	Dspace	Alumnos, Profesores, Investigadores	Tesis	Handle/1026	CURP	Temas CONACYT	UNESCO / LEM / INE / OIT	BY-NC-ND	Formato	Si	https://flacso.repositorioinstitucional.mx/jspui/	Busqueda Simple y Avanzada	gob.mx
15	Corporación Mexicana de Investigación en Materiales S.A. de C.V.	Repositorio Institucional de COMMSA	RIE	Lit	Dspace	Alumnos, Profesores, Investigadores	Publicaciones y Tesis	Handle/1022	CURP	Temas CONACYT	No	BY-NC-ND	Formato	Si	https://commsa.repositorioinstitucional.mx/jspui/	Busqueda Simple y Avanzada	gob.mx
16	Centro de Investigación en Materiales Avanzados, S.C.	Fuente de Objetos Científicos Open Access: Repositorio Institucional de CIMAV	FOCO	Lit	Dspace	Alumnos, Profesores, Investigadores	Memoria de Congreso, TesisMa, Artículo,	Handle/1004	CURP	Temas CONACYT	No	ABOUT/CCO/	Formato	Si	https://cimav.repositorioinstitucional.mx/jspui/	Busqueda Simple y Avanzada	gob.mx
17	Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California	Repositorio Institucional CICESE	RIC	Lit	Dspace	Alumnos, Profesores, Investigadores	TesisMa, Memorias de Congreso, Artículos	Handle/1007	CURP	Temas CONACYT	No	BY-NC-ND	Área	Si	https://cicese.repositorioinstitucional.mx/jspui/	Busqueda Simple y Avanzada	gob.mx
18	Centro de Ingeniería y Desarrollo Industrial	Repositorio Institucional de CIDESI	CIDESIRepositorio	Lit	Dspace	Alumnos, Profesores, Investigadores	Memorias de Congreso, TesisMa, TesisDo	Handle/1024	CURP	Temas CONACYT	Tesaurus propio	BY-NC-ND	Área / Tema	Si	https://cidesi.repositorioinstitucional.mx/jspui/	Busqueda Simple y Avanzada	gob.mx
19	Centro de Investigación y Desarrollo Tecnológico en Electroquímica.	Repositorio Institucional CIDETEQ	RI-CIDETEQ	Lit	Dspace	Todo Tipo de Usuarios	Proyectos y Tesis	Handle/1021	CURP	Temas CONACYT	Tesaurus propio	BY-NC-ND	Área / Formato	Si	https://cideteq.repositorioinstitucional.mx/jspui/	Busqueda Simple y Avanzada	gob.mx
20	Centro de Investigaciones Biológicas del Noroeste, S.C.	Repositorio Institucional CIBNOR	CIBNOR	Lit	Dspace	Alumnos, Profesores, Investigadores	Tesis, Artículos, Libros	Handle/1001	CURP	Temas CONACYT	Tesaurus propio	BY-NC-ND	Formato	Si	https://cibnor.repositorioinstitucional.mx/jspui/	Busqueda Simple y Avanzada	gob.mx
21	El Colegio de Michoacán, A.C.	Repositorio Institucional de COLMICH	ricolmich	Lit	Dspace	Alumnos, Profesores, Investigadores	Artículos, TesisMa, TesisDo, Libro, Capítulo de Libro	Handle/1016	ORCID	Temas CONACYT	LEMB	BY-NC-ND	Área / Formato	Si	https://colmich.repositorioinstitucional.mx/jspui/	Busqueda Simple y Avanzada	gob.mx
22	Centro de Investigación en Geografía y Geomática Ing. Jorge L. Tamayo	Repositorio Institucional de CENTROGEO	CENTROGEO	Lit	Dspace	GeneralPublic	Memoria de Congreso, Artículos	Handle/1012	ORCID	Temas CONACYT	No	BY-NC-ND/	Área	Si	https://centrogeo.repositorioinstitucional.mx/jspui/	Busqueda Simple y Avanzada	gob.mx
23	Centro de Investigaciones en Óptica, A.C.	Repositorio Institucional del CIO	REPCIO	Lit	Dspace	Alumnos, Profesores, Investigadores	Tesis y Artículos, Poster, Memorias de Congreso	Handle/1002	CURP	Temas CONACYT	No	BY-NC-ND	Formato	Si	https://cio.repositorioinstitucional.mx/jspui/	Busqueda Simple y Avanzada	gob.mx
24	Centro de Investigación en Matemáticas Ac	Repositorio Institucional CIMAT	CIMAT	Lit	Dspace	Alumnos, Profesores, Investigadores	Tesis, Artículos, Libros y Capítulo de Libro	Handle/1008	ORCID	Temas CONACYT	No	BY-NC	Área / Formato	Si	https://imat.repositorioinstitucional.mx/jspui/	Busqueda Simple y Avanzada	gob.mx
25	El Colegio de la Frontera Sur	Repositorio Institucional de ECOSUR	RIE	Lit	Dspace	Alumnos, Profesores, Investigadores	Tesis, Artículos, Capítulo de Libros y Tesis	Handle/1017	ORCID	Temas CONACYT	Tesaurus propio	BY-NC-ND	Formato	Si	https://ecosur.repositorioinstitucional.mx/jspui/	Busqueda Simple y Avanzada	gob.mx
26	Centro de Investigación en Alimentación y Desarrollo, A.C.	Repositorio Institucional del CIAD, A.C.	CIAD	Lit	Dspace	Librarians, Companies, Students, Communitygroups, Researchers, Teachers	Artículos, Libros y Tesis	Handle/1006	CURP	Temas CONACYT	Tesaurus propio	BY-NC-ND	Formato	Si	https://ciad.repositorioinstitucional.mx/jspui/	Busqueda Simple y Avanzada	gob.mx
27	El Colegio de la Frontera Norte A. C.	Repositorio Institucional de El Colegio de la Frontera Norte	RIEcolef	Lit	Dspace	Alumnos, Profesores, Investigadores	Artículos, Libros y Tesis	Handle/1014	CURP	Temas CONACYT	Tesaurus propio	BY-ND	Formato	Si	https://colef.repositorioinstitucional.mx/jspui/	Busqueda Simple y Avanzada	gob.mx
28	CIATEQ, A. C.	CIATEQ Digital	CIATEQ	Lit	Dspace	GeneralPublic	Documentos, Tesis y Propiedad Intelectual	Handle/1020	CURP	Temas CONACYT	Tesaurus propio	LICENSES/BI/	Área / Formato	Si	https://ciateq.repositorioinstitucional.mx/jspui/	Busqueda Simple y Avanzada	gob.mx
29	Universidad Autónoma de San Luis Potosí	Repositorio Institucional NINIVE	NINIVE	Lit	Dspace	Alumnos, Profesores, Investigadores	Documentos, Publicaciones, Materiales Audiovisuales, Archivo Institucional, Patrimonio documental, Recursos Educativos	Handle/	CURP	Temas CONACYT	No	BY-NC-ND / BY-NC-ND/2.5/MI/	Área / Formato	Si	https://ninive.uslp.mx/xmlui/	Sólo Busqueda Simple	XMLUI Default
30	Instituto Nacional de Medicina Genómica	Repositorio del Instituto Nacional de Medicina Genómica	inmegen	Lit	Dspace	Alumnos, Profesores, Investigadores	Artículos	123456789	CURP	Temas CONACYT	No	BY-NC/4.0	Formato	No	http://repositorio.inmegen.gob.mx/	Busqueda Simple	JSPUI personalizada
31	Universidad de Colima	Repositorio Institucional de la Universidad de Colima	BIBLOS	Lit	Dspace	Alumnos, Profesores, Investigadores	Producción Científica Publicaciones, Patente, Tesis, Memorias, Posters	Handle/123456789/	NO	Temas CONACYT	Tesaurus propio	LICENSES/BI/4.0	Área	Si	http://biblos.uco.mx/	Busqueda Por Autor y Título	JSPUI personalizada
32	Universidad Autónoma de Ciudad Juárez	Repositorio Institucional de la Universidad Autónoma de Ciudad Juárez	RI	Datos	VuFind	Academicos	Programas (Eventos de Academicos) Publicaciones y Multimedia	/20.500.11961/	NO	Temas CONACYT	No	CCO	Área / Formato	Si	http://ri.uaq.mx/vufind/	Busqueda por Tópicos, Tipo (Formato), Coleccion	Personalizada
33	Universidad Autónoma de Tamaulipas	Repositorio Institucional de la Universidad Autónoma de Tamaulipas	RIUAT	Lit	Dspace	GeneralPublic	Publicaciones y Recursos Educativos (Multimedia)	Handle/123456789/	CVU	Temas CONACYT	No	BY NC ND	Área	Si	http://riuat.uat.edu.mx/	Busqueda Simple y Avanzada	JSPUI personalizada
34	Instituto Tecnológico Superior de Centla	Repositorio Institucional del Instituto Tecnológico Superior de CENTLA	RIITSCe	Lit	Dspace	Alumnos, Profesores, Investigadores	Artículos, TesisMa, Libros y Capítulos de Libro	Handle/123456789/	NO	Temas CONACYT	No	BY/4.0	Área	No	http://repositorioitscce.iscentla.edu.mx/riitce/	Busqueda Simple y Avanzada Por Autor, Materia, Nivel de Acceso, Editor	JSPUI personalizada
35	Instituto Mexicano de Tecnología del Agua	Repositorio Institucional del IMTA	RIMTA	Lit	Dspace	Alumnos, Profesores, Investigadores	Artículo, Documento de trabajo, Libro, Objeto de conferencia, Capítulos de Libros	Hdl.Handle.Net/20.500.12013/1.311; No Sirve	RN	Temas CONACYT	Tesaurus propio	BY-NC-ND/4.0	Temas / Formato	No	http://repositorio.imta.mx/	Busqueda Simple	XMLUI personalizada
36	Universidad Autónoma de Nayarit	Repositorio Institucional ARAMARA	AramaraUAN	Lit	Dspace	Público En General		Handle/123456789/216	NO	Temas CONACYT	No	BY-NC/4.0	Área / Temática / Administrativa	Si	http://dspace.uan.mx:8080/	Busqueda Simple	JSPUI personalizada
37	Instituto Nacional de Pediatría	Repositorio del Instituto Nacional de Pediatría	RINP	Lit	Dspace	Especialistas	Artículos y Tesis	Handle/20.500.12103/3157	CURP	Temas CONACYT	Tesaurus propio	BY (ACTA PEDIATRICA) /4.0; BY-NC-ND	Área / Formato	Si	http://repositorio.pediatria.gob.mx:8180/	Busqueda Simple	JSPUI Default
38	Universidad de Guanajuato	Repositorio Institucional de la Universidad de Guanajuato	RIUGTO	Lit	Dspace	Especialistas	Artículos	Handle/20.500.12059/	CURP	Temas CONACYT	No	BY-NC-ND	Área / Temática	Si	http://www.repositorio.ugto.mx/	Busqueda Simple	JSPUI personalizada
39	Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz	Repositorio del Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz	RINPFM	Lit	Dspace	Especialistas	Artículos	Handle/123456789	CURP	Temas CONACYT	No	BY-NC-SA	Formato	Si	http://repositorio.inprf.gob.mx/	Busqueda Simple Autor, Título y Materia	XMLUI personalizada
40	Universidad Autónoma de Zacatecas	Repositorio Institucional CAXCÁN	RIcaxcan	Lit	Dspace	Alumnos, Profesores, Investigadores	Documentos, Tesis	2.050.011.845	ORCID	Temas CONACYT	Tesaurus propio	BY-NC-ND	Área	Si	http://ricaxcan.uaz.edu.mx/jspui/	Busqueda Simple	JSPUI personalizada
41	Universidad Autónoma de Nuevo León	Repositorio de Información Académica, Científica Tecnológica y de Innovación RIACTiS UANL	RIACTiS	Lit	EPrints 3	Alumnos, Profesores, Investigadores	Tesis, Artículos y Documentos	No	CURP	Temas CONACYT	No	BY-NC-ND/4.0	Área	No	http://riactis.uanl.mx/	Busqueda Simple y Avanzada Por Tema, Tipo de Documento, Departamento, Títulos, Autor	Personalizada
42	Universidad Autónoma Metropolitana	Repositorio Institucional Zolamati	RIZUAMA	Lit	Dspace	Estudiantes E Investigadores	TesisMa, TesisDo, Libro, Capítulo de Libro, Artículo	Http://Hdl.Handle.Net/11191	ORCID	Temas CONACYT	No	BY-NC-ND/4.0	Área	Si	http://zolamati.azc.uam.mx	Busqueda Simple	JSPUI personalizada
43	El Colegio de México	Colecciones Digitales COLMEX	Repositorio	Lit	?	Estudiantes E Investigadores	Tesis, Artículos y Libros	No	?	Temas CONACYT	No	BY/4.0 / BY-NC-ND/4.0	Formato	No	http://repositorio.colmex.mx/	Busqueda Simple	Personalizada
44	Universidad Juárez Autónoma de Tabasco	Repositorio Institucional de la Universidad Juárez Autónoma de Tabasco	RI-UJAT	Lit	Dspace	Estudiantes E Investigadores	Artículos, Tesis y Libro	Handle/20.500.12107	CVU	Temas CONACYT	No	BY-NC-ND; BY-NC (Artículos)	Área	Si	http://ri.ujat.mx/	Busqueda Simple	JSPUI personalizada
45	Instituto Nacional de Salud Pública	Repositorio Institucional Abierto de Conocimiento en Salud Pública	RIACSP	Lit	Dspace	Estudiantes E Investigadores	Tesis de Maestría y Doctorado y Artículos	Handle/20.500.12096	CURP	Temas CONACYT	No	BY-NC-ND/4.0	Formato	Si	http://repositorio.insp.mx:8080/jspui	Busqueda Simple	gob.mx
46	Universidad Autónoma del Estado de Morelos	Repositorio Institucional de Acceso Abierto de la Universidad Autónoma del Estado de Morelos	RIAA-UAEM	Lit	Dspace	Researchers	Artículos, Libro, Tesis D y Tesis M.	Handle/20.500.12055	CVU	Temas CONACYT	No	BY-NC-SA; BY-NC-ND (LIBRO)	Formato	Si	http://riaa.uaem.mx/XMLUI/	Busqueda Simple	XMLUI personalizada
47	Centro de Ciencias de la Atmósfera de la UNAM	Repositorio Institucional del Centro de Ciencias de la Atmósfera de la UNAM	RICCAUNAM	Lit	Datos	Geonetwork 3.0.4.0	Researchers	No	ORCID	Temas CONACYT	No	BY-NC-SA	Formato / Datos	No	http://ri.atmosfera.unam.mx/	Busqueda Simple	Personalizada
48	Tecnológico de Monterrey	Repositorio Institucional del Tecnológico de Monterrey	RIITEC	Lit	Dspace	Estudiantes E Investigadores	TesisMa, TesisDo y Artículos	11285/567725	?	Temas CONACYT	No	BY-NC-ND	Área / Formato	Si	https://repositorio.itesm.mx/	Busqueda Simple	XMLUI personalizada
49	Instituto de Investigaciones Sobre la Universidad y la Educación, UNAM	Repositorio Institucional del Instituto de Investigaciones Sobre la Universidad y la Educación	RI-ISUE	Lit	Dspace	Estudiantes E Investigadores	Artículos y Libros	Handle/issue_UNAM/	?	Temas CONACYT	No	BY-NC-ND	Formato	No	http://132.248.192.241:8080/xmlui/	Busqueda Simple Autor, Título y Materia	XMLUI personalizada
50	Universidad Pedagógica Nacional	XPLORA	RI-UPN	Lit	Dspace	Estudiantes E Investigadores	Tesis, Artículos, Libro, Capítulo de Libro	Handle/123456789/	CURP	Temas CONACYT	No	BY NC ND	Formato	Si	http://200.23.113.59:8080/jspui/	Busqueda Simple Por Título	JSPUI personalizada
51	Universidad Autónoma del Estado de México	Repositorio Institucional de la UAEM	RIUAEMEX	Lit	Dspace	Estudiantes E Investigadores	Publicaciones (TesisMa, TesisDo, Artículos, Libros y Capítul de Libro)	20.500.11799/6619	?	Temas CONACYT	?	BY NC ND	Área / Temática / Formato	Si	http://ri.uaemex.mx/	Busqueda Avanzada Título, Vía, Ruta, Materia, Auto, Fecha	JSPUI personalizada
52	Universidad Autónoma Metropolitana, Unidad Cuajimalpa	Concentración de Recursos de Información Científica y Académica, UAM Cuajimalpa	CONCENTRIC@	Lit	Dspace	Estudiantes E Investigadores	Artículos, Libros, Tesis y Multimedia	Handle/123456789	CVU	Temas CONACYT	No	BY-NC-ND	Formato	Si	http://ilita.cua.uam.mx:8080/jspui/	Busqueda Simple Autor, Título y Materia	JSPUI personalizada
53	Universidad Autónoma Agraria Antonio Narro	Repositorio Digital CID-UAAAN	R.D.C.I.D.-U.A.A.A.N.	Lit	Dspace	Estudiantes E Investigadores	Artículos	Handle/123456789	?	Temas CONACYT	No	BY-NC-ND; CCO Artículos; CC BY-NC-ND	Área	Si	http://repositorio.uaaan.mx:8080/xmlui/	Busqueda Simple	XMLUI personalizada
54	Centro de Investigaciones Sobre América del Norte - UNAM	Memoria Institucional CISAN, Repositorio Institucional, UNAM	MICISAN	Lit	Dspace	Estudiantes E Investigadores	Artículos	Handle/123456789	?	Temas CONACYT	No	BY-NC-ND	Área	Si	http://ru.micisan.unam.mx:8080/XMLUI	Busqueda Simple	XMLUI personalizada
55	Dirección General de Cómputo y de Tecnología de la Información y Comunicación	Repositorio Universitario de la Dirección General de Cómputo y de Tecnología de la Información y Comunicación	RU-TIC	Lit	Dspace	Estudiantes E Investigadores	Artículos, Libros, Tesis y Multimedia	Handle/123456789	NO	Temas CONACYT	No	CC BY-NC	Área	Si	http://ru.tic.unam.mx/tic/	Busqueda Simple, Busqueda Avanzada y Filas Avanzadas	XMLUI personalizada
56	Universidad Politécnica de Puebla	RI-UPPUE: Repositorio Institucional de la Universidad Politécnica de Puebla	RI-UPPue	Lit	Dspace	Estudiantes E Investigadores	Artículos	Handle/123456789	CVU	Temas CONACYT	No	BY NC ND; BY-NC-ND	Área / Formato	Si	http://repositorio.uppuebla.edu.mx:8080/xmlui/	Busqueda Simple, Busqueda Avanzada y Filas Avanzadas	XMLUI personalizada
57	Universidad Panamericana	Repositorio Institucional SCRIPTA de la Universidad Panamericana	SCRIPTA	Lit	Dspace	Estudiantes E Investigadores	Artículos	Handle/123456789	?	Temas CONACYT	No	BY-NC-ND/4.0	Áreas / Formato	No	http://scripta.up.edu.mx/xmlui/	Busqueda Simple, Busqueda Avanzada	XMLUI personalizada
58	Universidad Autónoma de Baja California Sur	Repositorio Electrónico Institucional de la Universidad Autónoma de Baja California Sur	RI/UABCS	Lit	Dspace	Estudiantes E Investigadores	Tesis, Tesis	Handle/23080/	?	Temas CONACYT	No	BY-NC-ND/4.0	Área	?	http://rep.uabcs.mx/	?	?
59	Universidad Autónoma Metropolitana Unidad Lerma	XOGI Repositorio Institucional UAM Lerma	xogi	Lit	Dspace	Alumnos, Profesores, Investigadores	Congreso, Artículos, Libros y Capítulos de Libro	Http://Hdl.Handle.Net/20.500									

ANEXO Núm. 1
Cuadro comparativo de los Repositorios Institucionales que interoperan con el Repositorio Nacional hasta el 30 de junio de 2019

NUM.	INSTITUCIÓN	NOMBRE DEL RI	ACRÓNIMO	TIPO DE RI LITERATURA - DATOS	SOFTWARE	COMUNIDAD DIRIGIDA	TIPOLOGÍA DE OBJETOS DIGITALES	IDENTIFICADOR DEL RECURSO	IDENTIFICADOR PERSONAL	TEMAS CONACYT	ÁREAS TEMÁTICAS PROPIA	LICENCIAS	ORGANIZACIÓN COMUNIDADES	ACCESO AL OAI	URL OFICIAL	BUSQUEDA	TIPO DE INTERFAZ
63	Universidad Iberoamericana Ciudad de México	Repositorio Institucional Universidad Iberoamericana Ciudad de México	RI-LIBERO	Lit	Dspace	Alumnos, Profesores, Investigadores	Artículos, Tesi, Tesd	Handle/ibero/	?	Temas CONACYT	No	BY-NC-ND	Temática	Si	http://ri.libero.mx/	Busqueda Simple	JSPUI personalizada
64	Universidad de Sonora	Repositorio Institucional UNISON	RI-UNISON	Lit	Dspace	Alumnos, Profesores, Investigadores	TesisMa, TesisDo	Handle/Unison	CVU ; CURP	Temas CONACYT	No	BY-NC-ND	Formato	Si	http://www.repositorioinstitucional.uson.mx	Busqueda Simple	XMLUI personalizada
65	Escuela Nacional de Estudios Superiores Unidad Morelia - Laboratorio Nacional de Análisis y Síntesis Ecológica	REDAEC - Repositorio de Datos Ecológicos del Laboratorio Nacional de Análisis y Síntesis Ecológica (LANASE)	REDAEC	Lit	Dataverse	Investigadores	Datos	No, sólo el número de especies	?	Temas CONACYT	No	LICENSES/0/4.0/	Área / Dataset Taxonomía	Si	http://www.redaec.unam.mx/index	Busqueda Por Filtros	Personalizada
66	Instituto de Química - UNAM	Repositorio del Instituto de Química	RIQ	Lit	Dspace	Alumnos, Profesores, Investigadores	Artículo, TesisMa, TesisDo y Memorias	Handle/20.500.12114/565	CVU/ORCID	Temas CONACYT	No	BY-NC-ND/4.0	Área / Temática	Si	http://rdu.iquimica.unam.mx/	Busqueda Simple	JSPUI Default
67	Universidad de Las Américas Puebla	Repositorio Institucional UDLAP: POHUA	RI-UDLAP	Lit	Dspace	Alumnos, Profesores, Investigadores	TesisMa, TesisDo y Artículo	Handle/123456789/	CVU	Temas CONACYT	No	BY-NC-ND/4.0	Formato	Si	http://repositorio.udlap.mx/	Busqueda Simple, Avanzada y Filtros	XMLUI personalizada
68	Instituto de Investigaciones Bibliotecológicas y de la Información	Repositorio Institucional del Instituto de Investigaciones Bibliotecológicas y de la Información de la Universidad Nacional Autónoma de México	Repositorio_IIIBI_UNA M	Lit	Dspace	Alumnos, Profesores, Investigadores	Capítulo de Libro, Artículo, TesisMa, Memoria de Congreso y Libros	Handle/ibibi_UNAM/	ORCID	Temas CONACYT	?	BY-NC-SA (LIBRO) ; BY-NC-ND	Formato	No	http://ru.ibibi.unam.mx/	Busqueda Simple, Avanzada	JSPUI personalizada
69	Centro Internacional de Mejoramiento de Maíz y Trigo	Repositorio Institucional de Datos y Software de Investigación del CIMMYT	CIMMYT-Datos-SW	Lit	Dataverse	Investigadores	Naturaleza del dato	hdl:11259/10548183	ORCID	FAO / NBC	FAO / NBC	BY/4.0	Datos Tipología	Si	https://data.cimmyt.org/dataverse/root	Especializada para Dataset	Personalizada
70	Centro Internacional de Mejoramiento de Maíz y Trigo	Repositorio Institucional de Publicaciones Multimedia del CIMMYT	CIMMYT-Multimedia	Lit	Dspace	Investigadores	Documentos, Informes Anuales, Libros, Artículos de Revistas y Recursos Multimedia, Artículo, Libro, Poster de Congreso, Documento de Trabajo	Handle/10884	ORCID	Temas CONACYT	FAO	BY-NC	Área		https://repository.cimmyt.org/	Personalizada, Serie, Autor, Fecha, Tipo, Materia, Región, País	Personalizada
71	Universidad Autónoma de Chiapas	Repositorio Institucional de la Universidad Autónoma de Chiapas	RIUNACH	Lit	Dspace	Alumnos, Profesores, Investigadores	TesisMa, TesisDo	Handle/123456789	CURP	Temas CONACYT	No	BY-NC-ND	Área Administrativa / Formato	Si	http://www.repositorio.unach.mx/jspui/	Busqueda simple	No Acceso
72	Servicio Sismológico Nacional, Instituto de Geofísica, UNAM	Repositorio de Datos del Servicio Sismológico Nacional	SSNMx	Lit	Dspace	Investigadores, Profesores	Datos y Publicaciones	Handle/SSnmx/	CURP	Temas CONACYT	LCC / GEOFREF	BY-NC-ND/4.0	Datos Tipología / Formato	Si	http://rigeofisica.ssn.unam.mx/jspui/	Personalizada	JSPUI personalizada
73	Instituto Politécnico Nacional Unidad Sinaloa	Ciencia Sinaloa IPN	CSIN-IPN	Lit	Dspace	Alumnos, Profesores, Investigadores	TesisMa, TesisDo	Handle/123456789/		Temas CONACYT	No	ABOUT/CC0/	Formato	Si	http://www.cienciasinaloa.ipn.mx/80/jspui/	Personalizada	No Acceso
74	Universidad Autónoma de Yucatán	Repositorio Digital Institucional de la Universidad Autónoma de Yucatán	REDI-UADY	Lit	Dspace	Alumnos, Profesores, Investigadores	Artículos	Handle/123456789	CURP	Temas CONACYT	No	BY-NC-ND	Área / Temática	Si	http://redi.uady.mx:8080/	Personalizada	XMLUI personalizada
75	Instituto Nacional de Geriátria	Repositorio Institucional de INGER	R-GERIATRIA	Lit	Dspace	Alumnos, Profesores, Investigadores	Artículos	Handle/20.500.12100/	CURP	Temas CONACYT	No	LICENSES/0/4.0/ ; BY-NC-ND/	Formato	No	http://189.209.180.190/jspui/	busqueda simple y avanzada	gob.mx
76	Unidad Profesional Interdisciplinaria de Ingeniería y Ciencias Sociales y Administrativas	Repositorio UPIICSA IPN	REPOUICSA	Lit	Dspace	Alumnos, Profesores, Investigadores	Artículos de Revistas, Libros, Tesis de Licenciatura, Manual, Recurso Educativo, Evento Académico, Material Audiovisual, Capítulos de Libros y TesisMa	Handle/20.500.12271/	?	Temas CONACYT	No	BY-NC-ND	Área	No	http://repositorio.upicisa.ipn.mx/	Personalizada	XMLUI personalizada
77	Universidad Autónoma Metropolitana Unidad Iztapalapa	Repositorio Institucional de la Universidad Autónoma Metropolitana Unidad Iztapalapa	BINDANI	Lit	Dspace	Alumnos, Profesores, Investigadores	TesisMa, TesisDo	Handle/11191/	?	Temas CONACYT	No	BY-NC-ND	Área	Si	http://zaomati.azc.uam.mx/	Personalizada	JSPUI personalizada
78	Universidad Nacional Autónoma de México, Instituto de Ciencias del Mar y Limnología	Repositorio UNINMAR del Instituto de Ciencias del Mar y Limnología	UNINMAR	Datos	Dspace	Investigadores, Profesores	Dataset de la Unidad de Informática Marina	Handle/20.500.12201/	ORCID	Temas CONACYT	No	BY-NC/4.0	Área	No	http://metadata.lcmly.unam.mx/	Avanzada	XMLUI personalizada
79	Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, UNAM	Repositorio Institucional del Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas	RIIMAS	Lit	Dspace	Alumnos, Profesores, Investigadores	Artículo, Capítulo de Libro	Handle/123456789/	ORCID	Temas CONACYT	No	BY-NC-ND ; BY ; BY	Formato	Si	http://www.ru.limas.unam.mx/	busqueda simple y avanzada	XMLUI Default
80	Centro de Ciencias Genómicas	Repositorio Institucional de Conocimiento Genómico	RICG	Datos	?	Investigadores, Profesores	Dataset	20.500.12148/Mb6hag	?	Temas CONACYT	?	BY-NC-SA	Datos Familias, Taxonomías	?	20.500.12148/N8K1QJ	?	No acceso
81	Universidad Tecnológica del Centro de Veracruz	Universidad Tecnológica del Centro de Veracruz	Rein-UTCV	Lit			Preimpreso, Poster de Congreso, Artículo, Memoria de Congreso, Conferencia	Handle/123456789/				BY ;	Formato		http://rein.utcv.edu.mx/	busqueda simple y avanzada	JSPUI personalizada
82	Instituto de Geofísica (Michoacán) UNAM	Repositorio de Recursos Geotérmicos Para México	geotermx	Lit	Dspace	Investigadores, Profesores	Dataset e Informes Técnicos	Handle/Geotermx/	ORCID	Temas CONACYT	LCC / GEOFREF	BY-NC-ND	Datos tipologías y gps		http://rigeotermia.geofisica.unam.mx/	Manifestacion, Año y Título	JSPUI personalizada
83	Instituto Politécnico Nacional	Repositorio CITEDI IPN	CITEDI	Lit	Dspace	Alumnos, Profesores, Investigadores	TesisMa, Memoria de Congreso, TesisDo, Artículo y Otro (Trabajo Terminal)	Handle/123456789/	CVU	Temas CONACYT	No	BY-NC-ND	Área	Si	https://www.repositoriodigital.ipn.mx/	Personalizada XMLUI	JSPUI personalizada
84	Facultad de Estudios Superiores Iztacala	Repositorio Universitario de Prosa Científica y de Divulgación de la Facultad de Estudios Superiores Iztacala	RU-Iztacala	Lit	Omeka	Alumnos, Profesores, Investigadores	Artículo, Libro, Capítulo de Libro, Objeto de Congreso	Items/Show/	CVU	Temas CONACYT	No	BY-NC-ND	Formato	No	https://ru.iztacala.unam.mx/	Busqueda simple	JSPUI personalizada
85	Dirección General de Divulgación de la Ciencia - UNAM	AMEYALLI - Repositorio Universitario Especializado en Comunicación Pública de la Ciencia	RUEPCP	Lit	Dspace	Alumnos, Profesores, Investigadores	Dataset, TesisMa, Artículo, TesisDo	Handle/123456789	CURP	Temas CONACYT	No	BY-NC-ND	Área	No	http://ru.ameyalli.dgdc.unam.mx/	Busqueda simple y filtros	JSPUI personalizada
86	Dirección General de Divulgación de la Ciencia - UNAM	AMEYALLI - Repositorio Universitario Especializado en Comunicación Pública de la Ciencia	RUEPCP	Lit	Dspace	Alumnos, Profesores, Investigadores	Dataset, TesisMa, Artículo, TesisDo	Handle/123456789	CURP	Temas CONACYT	No	BY-NC-ND	Área	No	http://ru.ameyalli.dgdc.unam.mx/	Busqueda simple y filtros	JSPUI personalizada
87	Instituto de Geofísica (Michoacán) UNAM RICE	Repositorio Institucional de Clima Espacial	RICE	Datos	Dspace	Investigadores, Profesores	Dataset	Handle/Rice	ORCID	Temas CONACYT	LCC / GEOFREF / UNESCO	BY-NC-SA	Datos tipologías y gps	Si	http://www.rice.unam.mx:8080/XMLUI/	Por filtros	XMLUI Default
88	Universidad Autónoma de Querétaro	Mejora y Adecuación del Repositorio Institucional de la Universidad Autónoma de Querétaro	RIUAQ	Lit	Dspace	Alumnos, Profesores, Investigadores	TesisMa, TesisDo	Handle/123456789		Temas CONACYT	No	BY-ND	Área		http://ri-eg.uaq.mx/	Busqueda simple y filtros	JSPUI personalizada
89	Instituto Politécnico Nacional	Repositorio de Acceso Abierto a la Información Científica, Tecnológica y de Innovación del Centro de Biotecnología Genómica del IPN: Doctorado en Ciencias en Biotecnología en Red del IPN	RDCBIPN	Lit	Dspace	Alumnos, Profesores, Investigadores	Artículos, Patente	/handle/20.500.12273	CVU	Temas CONACYT	No	BY-NC ; ABOUT/CC0 ; BY-NC	Área	Si	http://rdcb.cbipn.mx/	Por comunidades	JSPUI personalizada
90	Universidad Michoacana de San Nicolás de Hidalgo	Repositorio Institucional de la Universidad Michoacana de San Nicolás de Hidalgo	RIUMSNH	Lit	Dspace	Alumnos, profesores, investigadores	TesisMa, TesisDo, Licenciatura y Especialidad	handle/DGB_UMICH	?	?	?	by-nc-nd/4.0	Formato	No Acc	http://bibliotecavirtual.dgb.umich.mx:8083/jspui/	Por comunidades	JSPUI Default
91	Benedicta Universidad Autónoma de Puebla	Repositorio Institucional de Acceso Abierto RIAA-BUAP	RIAA-BUAP	Lit	Dspace	GeneralPublic	Libros, Revistas y Tesis	20.500.12055/182	CVU	Temas CONACYT	No	/by-nc-sa/	Formato	Si	http://riaa.uaem.mx/	Busqueda simple y filtros	XMLUI personalizada
92	Instituto de Fisiología Celular - UNAM	Repositorio del Instituto de Fisiología Celular	REPIFC	Lit	Dspace	?	Fotografías, Logos y material gráfico	handle/RII	?	?	?	CC BY-NC-SA	Formato	No Acc	http://www.ru.imagenes.unam.mx/	Busqueda avanzada y filtros	JSPUI personalizada
93	Laboratorio Nacional de Informática Avanzada, A. C.	Datos Abiertos en Repositorio LANIA	DAR-LANIA	Lit	Dspace	Alumnos, profesores, investigadores	Artículos de congreso, revista, Capítulo de libro, Libro, Patente, Tesis, Poster, Reporte Técnico	/publication-detail/handle/237	CVU	Temas CONACYT	No	by-nc-nd	Formato	Si	http://repositorio.lania.mx/	Busqueda general y comunidades	JSPUI personalizada
94	Universidad Autónoma de Guerrero	Repositorio Institucional de Ciencia Abierta de la Universidad Autónoma de Guerrero	RIUAGro	Lit	Dspace	Alumnos, profesores, investigadores	TesisMa, TesisDo, Artículo, Revista Tlamarí UAGro, Autodeposito de Artículos y Tesis	handle/uagro/1	?	Temas CONACYT	Tesoro CAB / UNESCO	by-nc-nd	Formato / Área	Si	http://ri.uagro.mx/	Busqueda general y comunidades	JSPUI personalizada
95	Instituto Politécnico Nacional	Repositorio Institucional de Literatura del IPN-CIDIR Unidad Oaxaca	RIL-IPN-CIDIROAX	Lit	Dspace	Alumnos, profesores, investigadores	Master thesis (115) Doctoral thesis (14)	handle/LITER_CIDIROAX/30	?	Temas CONACYT	Tesoro propio	by-nc-nd/4.0	2 Formatos	No	http://literatura.cidiroaxaca.ipn.mx:8080/mlui/	Busqueda general y comunidades	XMLUI Default
96	Instituto Politécnico Nacional	Repositorio Institucional de Colecciones Científicas del IPN-CIDIR Unidad Oaxaca	RIC-IPN-CIDIROAX	Datos	Dspace	Profesores, investigadores	Familias y Tipologías (Material biológico)	handle/COLEC_CIDIROAX	?	Temas CONACYT	Tesoro propio	by-nc-nd/4.0	Datos tipologías y taxonomías	No	http://coleccion.cidiroaxaca.ipn.mx/jspui/	Busqueda por comunidades	JSPUI Default

** Nota las celdas marcadas de color rojo, son los Repositorios Institucionales que su grosor de son Datos.

ANEXO núm. 2

Ejemplo de una estructura de un archivo *DATASETNAMEreadme.txt*

Este archivo DATASETNAMEreadme.txt fue generado en YYYY-MM-DD por NAME <el texto de ayuda se incluye entre corchetes angulares y se puede eliminar antes de guardar>

INFORMACIÓN GENERAL

1. Título del conjunto de datos:
2. Información del autor
 - a. Información de contacto del investigador principal:
 - Nombre:
 - Institución:
 - Dirección:
 - Email:
 - b. Información de contacto del asociado o Co-investigador:
 - Nombre:
 - Institución:
 - Dirección:
 - Email:
 - c. Información de contacto alternativa:
 - Nombre:
 - Institución:
 - Dirección:
 - Email:
3. Fecha de recopilación de datos (fecha única, rango, fecha aproximada) <formato sugerido AAAA-MM-DD>:
4. Ubicación geográfica de la recopilación de datos <latitud, longitud o ciudad / región, estado, país, según corresponda>:
5. Información sobre fuentes de financiamiento que apoyaron la recopilación de datos:

COMPARTIR / ACCEDER A LA INFORMACIÓN

1. Licencias / restricciones impuestas a los datos:
2. Enlaces a publicaciones que citan o utilizan los datos.
3. Enlaces a otras ubicaciones de datos de acceso público:
4. Vínculos / relaciones con conjuntos de datos auxiliares:
5. ¿Los datos se obtuvieron de otra fuente? Sí / No
 - a. En afirmativo, indique la (s) fuente (s):
6. Cita recomendada para este conjunto de datos:

RESUMEN DE DATOS Y ARCHIVOS

1. Lista de archivos:

<enumere todos los archivos (o carpetas, según corresponda para la organización del conjunto de datos) contenidos en el conjunto de datos, con una breve descripción>
2. Relación entre archivos, si es importante:
3. Datos relacionados adicionales recopilados que no se incluyeron en el paquete de datos actual:
4. ¿Existen varias versiones del conjunto de datos? Sí / No
 - a. En caso afirmativo, nombre de los archivos que se actualizaron:
 - i. ¿Por qué se actualizó el archivo?
 - ii. ¿Cuándo se actualizó el archivo?

INFORMACION METODOLÓGICA

1. Descripción de los métodos utilizados para la recopilación / generación de datos:
<Incluir enlaces o referencias a publicaciones u otra documentación que contenga diseños experimentales o protocolos utilizados en la recopilación de datos>
2. Métodos para procesar los datos:
<describir cómo se generaron los datos enviados a partir de los datos sin procesar o recopilados>
3. Información específica del software o del instrumento necesaria para interpretar los datos:
<incluya el nombre completo y la versión del software, y cualquier paquete o biblioteca necesarios para ejecutar scripts>
4. Patrones e información de calibración, si procede:
5. Condiciones ambientales / experimentales:
6. Describa cualquier procedimiento de garantía de calidad realizado con los datos:
7. Personas involucradas en la recolección, procesamiento, análisis y / o envío de muestras:

INFORMACIÓN ESPECÍFICA DE DATOS PARA: [NOMBRE DE ARCHIVO]

<repita esta sección para cada conjunto de datos, carpeta o archivo, según corresponda>

1. Número de variables:
2. Número de casos / filas:
3. Lista de variables:
<enumere el (los) nombre (s) de variable, descripción (es), unidad (es) y etiquetas de valor según corresponda para cada>
4. Códigos de datos faltantes:
<código / símbolo de lista y definición>
5. Formatos especializados u otras abreviaturas utilizadas:

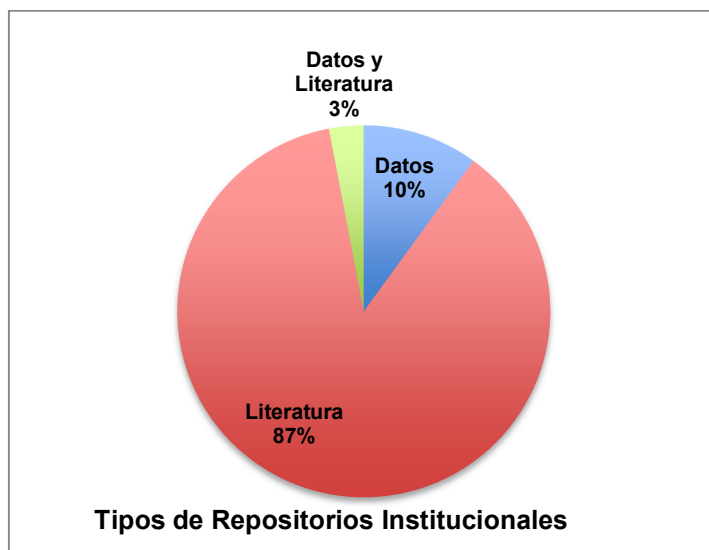
Tomado y trabajado en el *Taller de Organización de datos de investigación: como normalizar los nombres de los archivos, las versiones y establecimiento de formatos.*

Con el Instructor Lic. Luis Enrique Lescano Borrego y traducido por Minerva Castro Escamilla.

Anexo 3. Tipos de Repositorios Institucionales

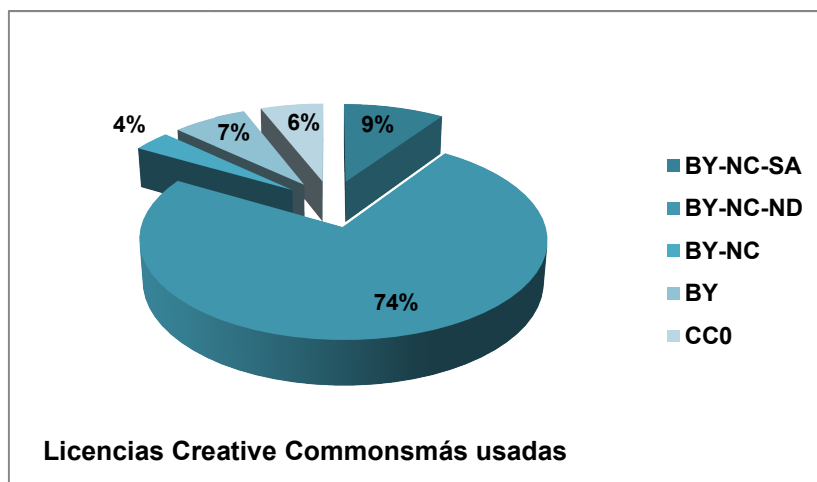
El Repositorio Nacional cuenta con 96 Repositorios Institucionales en los que se encuentran 3 tipos:

- a) Repositorio de Literatura
- b) Repositorio de Datos
- c) Repositorio de Datos y Literatura



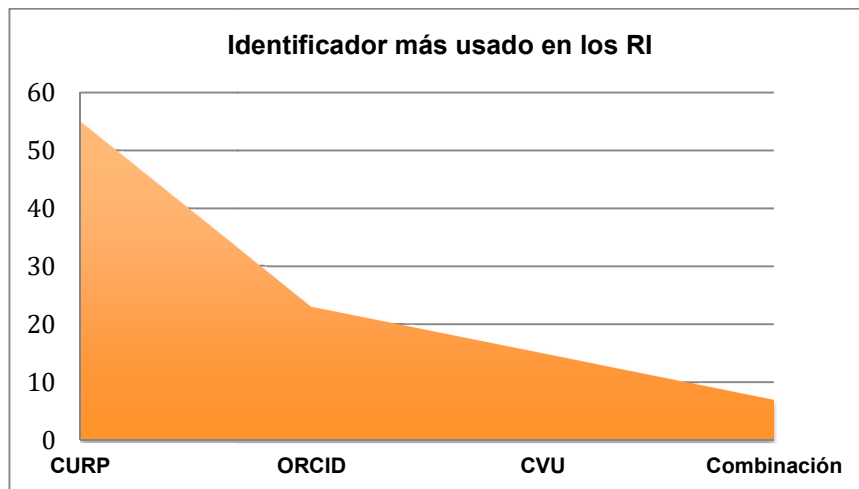
Anexo 4. Licenciamientos y establecimiento de permisos

En cuanto al resultado que arrojó la revisión de licenciamientos y asignación de permisos del uso de datos de las 6 licencias que existen, solo se utilizan 4 y se enlistan por orden de uso en cada uno de los RI. Es importante mencionar que también existen cartas con autorización de publicar el contenido en el RI, así como sesión de los derechos.



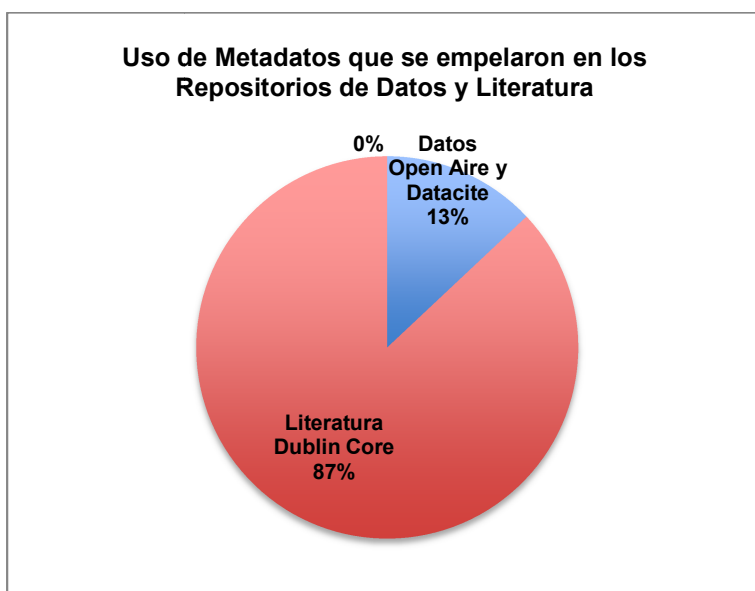
Anexo 5. Uso del tipo de identificador de autor

El Conacyt establece dos tipos de catálogos de servicios REST. Para los RI de Literatura cuenta con 20 servicios; y para RI de Datos, solo con 12. En ambos casos, deben interoperar con campos obligatorios. Se muestra el tipo de identificador de mayor uso en el campo de Autor cuando se lleva a cabo la conexión de los servicios REST con el Repositorio Nacional.



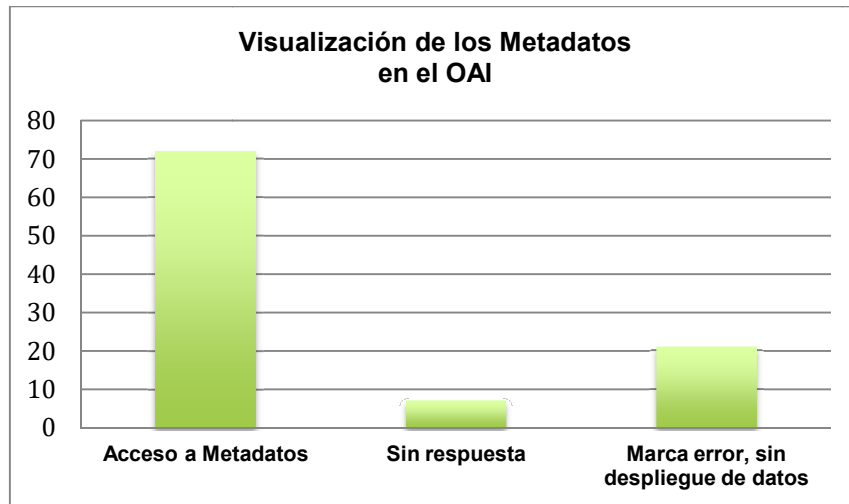
Anexo 6. Tipos de metadatos

Como lo marcan los *Lineamientos específicos para RI de Conacyt*, en los RI de Literatura utiliza Dublin Core simplificado y en los de Datos se usan las directrices de OpenAIRE 3.0, que complementan a DataCite y el resultado fue:



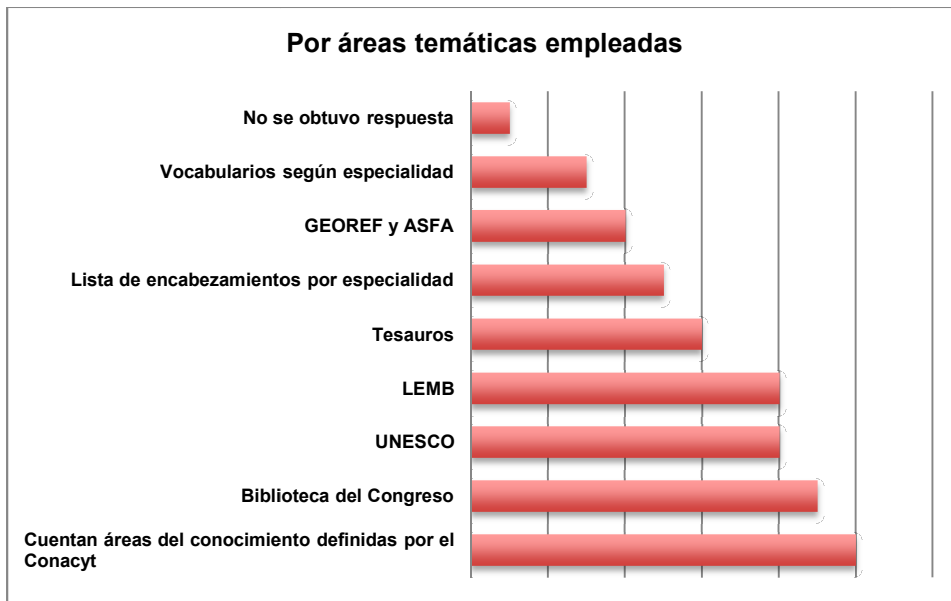
Anexo 7. Visualización de los metadatos en el OAI

Se accedió al URL de su OAI-PMH para ver la visualización de los Metadatos y el resultado fue el que se muestra en el gráfico:



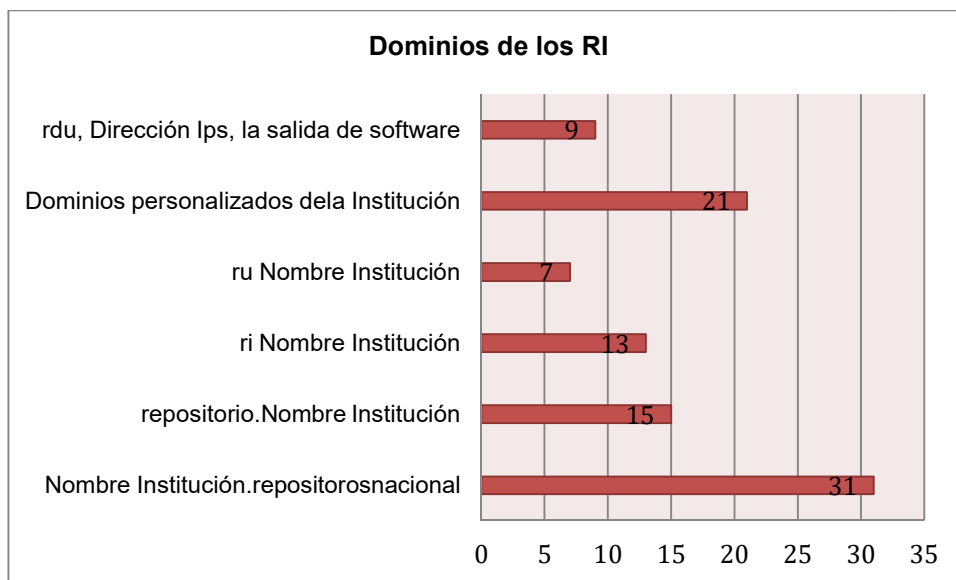
Anexo 8. Áreas temáticas empleadas en los RI

Dentro de sus Lineamientos, Conacyt solicita que se integren sus siete áreas del conocimiento dentro de sus metadatos. Pero además es posible por especialidad según RI, a fin de ser mejor indizadas. En este resultado, se emplearon las siguientes:



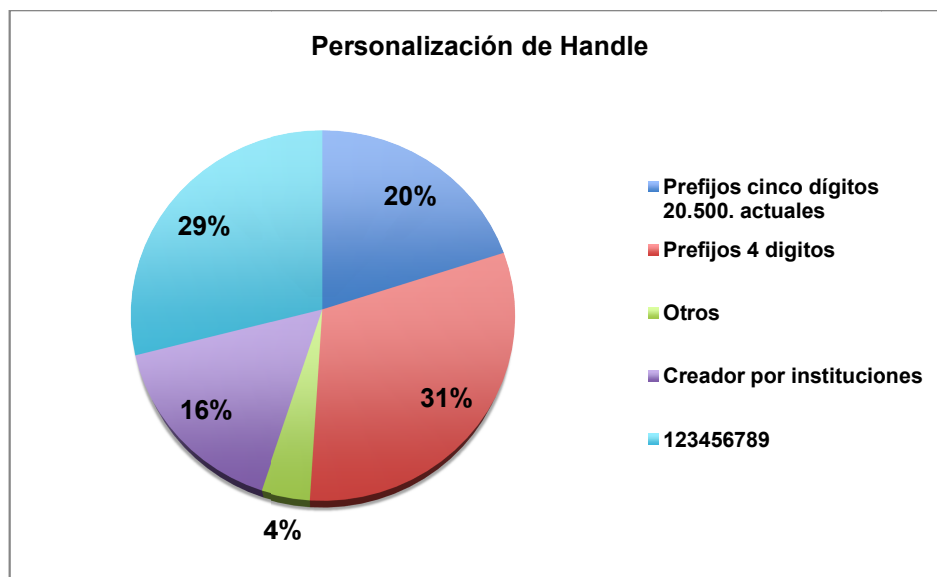
Anexo 9. Personalización de dominios URL

Dentro de la comunidad de los RI, existen diferentes tipos de dominios que utilizaron las instituciones para la localización del sitio web:



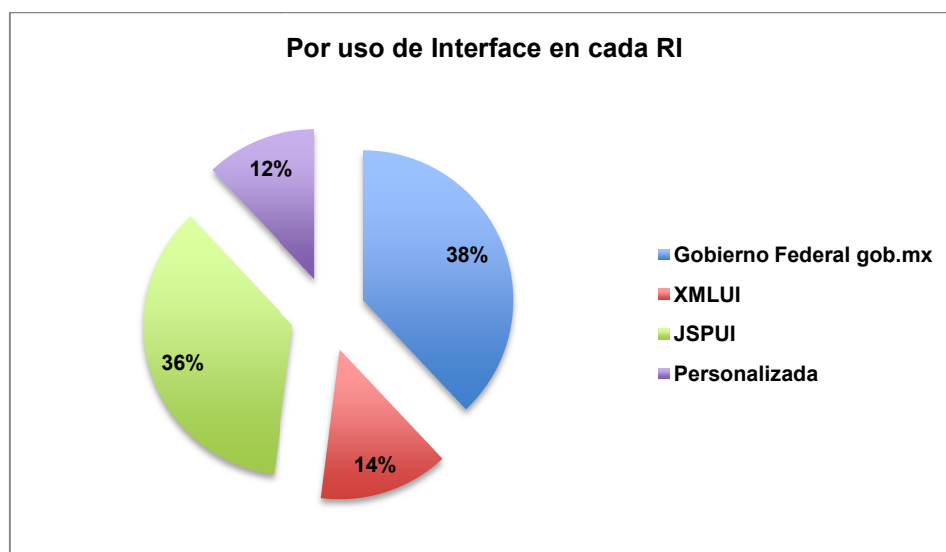
Anexo 10. Personalización de los objetos digitales HANDLE

Se muestra la variabilidad de tipos de Handle que se emplean en cada uno de los Repositorios de las Instituciones.



Anexo 11. Tipo de interface empleada en los RI

Muestra el tipo de interface que se utiliza en los RI, según las necesidades para presentar la información:



Anexo 12. Tipo de búsqueda en los Repositorios

Existen diversas estructuras de búsqueda que se emplearon en cada uno de los RI, tanto de Literatura como de Datos para recuperar la información:

