

La gobernanza de los datos de investigación en el contexto de su organización y sistematización

Eder Ávila-Barrientos*

Artículo recibido:
22 de febrero de 2023
Artículo aceptado:
2 de mayo de 2023
Artículo de investigación

RESUMEN

La *gobernanza* es un concepto aplicado al estudio de las ciencias políticas y sociales para explicar las regulaciones y buenas prácticas del Estado en temas de la ciudadanía y la administración pública. La gobernanza de los datos de investigación engloba el desarrollo de políticas para propiciar el efectivo manejo de los datos; como parte de este manejo, la sistematización y la organización son dos procesos que intervienen en la recuperación y accesibilidad a dichos datos. Por lo tanto, el objetivo de este trabajo consiste en identificar los atributos de la gobernanza de los datos de investigación, para comprender su interacción en el contexto de su organización y sistematización. A través de una metodología basada en dos etapas,

* Instituto de Investigaciones Bibliotecológicas y de la Información, Universidad Nacional Autónoma de México eder@iibi.unam.mx

la primera de ellas, de carácter documental, apoyada en la hermenéutica del discurso, y la segunda en el método analítico-sintético para describir y representar a los datos de investigación, se presentan los hallazgos obtenidos en este estudio. Se estima que la creación y adaptación de un perfil de metadatos diseñado ex profeso para este tipo de datos puede contribuir a establecer un mecanismo interoperable para describir y representar a los datos de investigación en un sistema disponible en el contexto digital.

Palabras clave: Gobernanza; Datos de Investigación; Organización; Sistematización

Governance of research data in the context of its organization and systematization

Eder Ávila-Barrientos

ABSTRACT

Governance is a concept applied to the study of political and social sciences to explain the regulations and good practices of the State in matters of citizenship and public administration. The governance of research data encompasses the development of policies to promote effective data management, as part of this management, systematization and organization are two processes involved in the recovery and accessibility of said data. Therefore, the objective of this paper is to identify the attributes of research data governance in order to understand their interaction in the context of their organization and systematization. Through a methodology based on two stages, the first of which is documentary based on the hermeneutics of discourse and the second on the analytic-synthetic method to describe and represent research data, the findings obtained in this study are presented. It is estimated that the creation and adaptation of a metadata profile designed specifically for this type of data can contribute to establishing an interoperable mechanism to describe and represent research data in a system available in the digital context.

Keywords: Governance; Research Data; Organization; Systematization

INTRODUCCIÓN

Desde la década de los noventa del siglo pasado, la *gobernanza* es un concepto que se ha utilizado desde una perspectiva política para referirse a la eficacia, calidad y efectiva dirección de la participación del Estado en asuntos de la ciudadanía. Este concepto ha sido adaptado y configurado en un contexto informativo en donde los datos son de suma relevancia para la toma de decisiones en el ámbito gubernamental.

En el contexto de la investigación científica y académica que se desarrolla con fondos públicos, la *gobernanza de los datos de investigación* (GDI) se refiere a las políticas y directrices que permiten regular el manejo y la gestión de dichos datos. Esto conlleva propiciar su acceso y reutilización mediante el uso de sistemas digitales.

En este sentido, la organización y sistematización de los datos de investigación son procesos que involucran el uso de estándares internacionales, como es el caso de esquemas de metadatos desarrollados ex profeso para llevar a cabo la descripción y la representación de los datos bajo un perfil común.

Bajo esta premisa, el presente trabajo pretende abordar los siguientes cuestionamientos:

- ¿De qué manera pueden describirse y representarse los datos de investigación para propiciar su gobernanza?
- ¿Cuál es la interacción entre la gobernanza de los datos y su descripción y representación?

De acuerdo con Koltay (2016: 304) “la expresión gobernanza de datos (GD) podría referirse a los órganos organizativos; reglas, políticas, estándares; derechos de decisión; asignación de responsabilidades y métodos de ejecución que hacen posible la gestión de los datos”. De esta manera, la gobernanza de datos engloba a los procesos y técnicas para propiciar el acceso a los datos en un marco común.

De esta manera, el objetivo que persigue este trabajo es identificar los atributos de la gobernanza de los datos de investigación y comprender su interacción en el contexto de su organización y sistematización. De este modo, el estudio de los datos de investigación requiere de una perspectiva holística que permita comprender las aristas que forman parte de su fenomenología.

Gobernanza de datos de investigación

En años recientes, la gobernanza de los datos ha cobrado notable relevancia, debido a su dinamismo y a su aparición en esferas y contextos que versan sobre dominios como la administración pública y la privada, el gobierno abierto, la gestión organizacional, la transparencia y la rendición de cuentas. Sin embargo, su corpus teórico y conceptual aún presenta patrones de incertidumbre, al no reflejar con claridad su epistemología y razón de ser.

Desde la óptica de la gestión de datos, la GDI permite identificar el uso de políticas y lineamientos para hacer accesibles los datos que forman parte de la administración, pública y privada, de las instituciones gubernamentales y de la ciencia abierta. De acuerdo con Weber *et al.* (2009: 6), la GDI pone de manifiesto un marco común para ejercer los derechos de decisión y las responsabilidades para fomentar el comportamiento deseable en el uso de los datos. Si bien esta definición puede prestarse a confusiones, son de resaltarse las buenas prácticas que se deben llevar a cabo para utilizar los datos en diversos contextos.

Por otra parte, la Organización para la Cooperación y el Desarrollo Económico (OECD, por sus siglas en inglés) manifiesta que la gobernanza de los datos se refiere a diversas estrategias, que incluyen disposiciones técnicas, normativas, reglamentarias e institucionales, que pueden afectar a los datos y su creación, recopilación, almacenamiento, uso, protección, acceso, intercambio y eliminación, todo esto incluido en los dominios de las políticas organizacionales y gubernamentales (OECD, 2022: 13).

De esta manera, las políticas nacionales para el acceso a los datos forman parte de un conjunto de instrumentos normativos que tienen el propósito de regular el manejo y la accesibilidad de los datos en el contexto público y gubernamental; estas políticas tienen una interacción con el uso de los datos en diversas actividades y contextos productivos. Aunado a ello, “la gobernanza de los datos contempla a los procesos estándares y modelos que en un entorno común describen integralmente quién puede tomar una serie de acciones sobre los datos, esto de acuerdo con métodos predeterminados y fundamentados en los derechos de acceso” (Kuiler, 2022: 286).

Dentro de estos métodos pueden ubicarse aquellos relacionados con una estructura de metadatos para organizar, representar y hacer accesibles a los datos en sus diferentes formas y naturaleza. Pues la GDI debe contemplar su efectiva organización y representación. La complejidad de los datos y su dinamismo refleja su carácter interdisciplinario, en donde cada vez se hace más frecuente el uso de la inteligencia artificial y de algoritmos para captar datos de diversas

fuentes. Sin embargo, esto trae consigo una serie de retos y consideraciones que, desde la óptica de la gobernanza, han sido discutidos por diversos autores.

Por ejemplo, Jansen *et al.* (2020: 7) han recabado e interpretado los principios de la gobernanza de los datos desde su evaluación, la cual permite detectar sus irregularidades, hasta establecer un juicio imparcial al momento de describirlos y representarlos. De acuerdo con Fleckenstein y Fellows (2018: 64), la gobernanza de datos se trata, en gran medida, de colaboración, es decir, de conectar a productores y consumidores de datos y determinar quién tomará las decisiones fundadas en el uso de dichos datos con base en su accesibilidad. Desde este enfoque, la GDI remite al desarrollo de buenas prácticas para gestionar los datos en diferentes entornos, dejando ver un proceso que consta de etapas específicas para manejarlos desde un punto de vista organizativo y apegado a su sistematización.

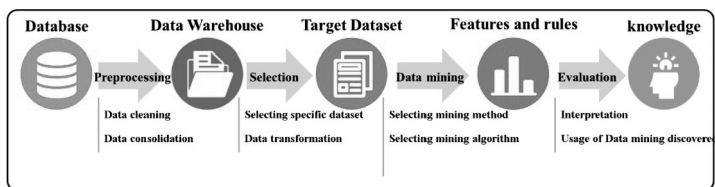


Figura 1. Four-steps data analytics process for data governance of NPO
Fuente: Wang *et al.* (2019, 500)

Dentro de la literatura especializada en el tema de GDI, se puede identificar una amplia gama de procesos que tratan de explicar la manera en cómo los datos pueden ser manejados para obtener de ellos variables de estudio. Estos procesos son acompañados, la mayoría de las veces, de una metodología *ex profeso*. Por ejemplo, en la *Figura 1* puede observarse el proceso de análisis para la gobernanza de datos que toman parte en organizaciones sin ánimo de lucro. En este proceso destacan fases que engloban un proceso general para dar tratamiento a los datos que están depositados en una base de datos o un silo en particular.

Desde la limpieza de los datos, hasta su acomodo y transformación, se trata de fases que tienen el propósito de dirigir el manejo de los datos a un asunto en particular, tomando en consideración las fases de análisis a las cuales se exponen los datos. Técnicas tales como la minería de datos y su procesamiento algorítmico contribuyen a una mejor interpretación de los datos, con el objetivo de construir conocimientos sobre un tema, un fenómeno o una problemática a solucionar.

Con el auge de la ciencia abierta y la necesidad de contar con datos de investigación para reutilizarlos en la generación de nuevos conocimientos, la GDI tiene una interacción con la generación de políticas que permitan dar solidez y apertura a los datos, sobre todo en entornos académicos y científicos. En este sentido, “la ciencia abierta es el movimiento que busca la apertura de la investigación

científica (métodos, instrumentos, datos, etc.) para beneficio de toda la sociedad. La ciencia abierta se presenta como un medio capaz de articular y dinamizar las políticas de ciencia, tecnología e innovación” (AmeliCA, 2019: párr. 1).

De esta manera, la ciencia abierta y la GDI interactúan en un entorno de investigación científica y académica que busca eliminar las barreras de acceso a la producción intelectual que se genera en estos contextos. Esto mediante el desarrollo de políticas que regulen dicha apertura.

En este sentido, “el concepto de ciencia abierta supone la transparencia de la metodología de investigación y observación, así como el amplio acceso a las colecciones de conjuntos de datos científicos utilizados y adquiridos como resultado de experimentos, la libre accesibilidad y reutilización de dichos datos, la disponibilidad y transparencia de las comunicaciones científicas y el uso de plataformas virtuales de apoyo al flujo de trabajo científico colaborativo” (Karmanskiy et al., 2016: 400).

Bajo esta premisa, en la *Figura 2* puede observarse la interacción entre las políticas, la ciencia abierta y la GDI tomando en cuenta el contexto de la investigación científica y académica en donde se establece un marco común que motiva la necesidad de organizar y sistematizar a los datos que se generan en el contexto académico y científico.

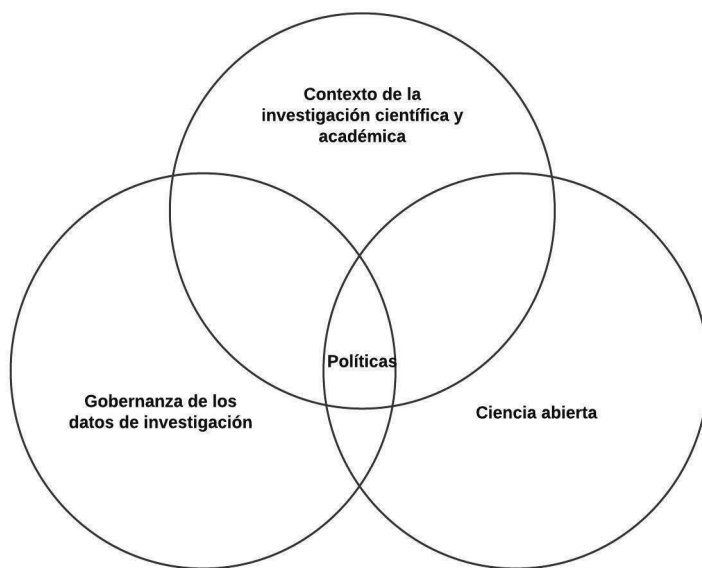


Figura 2. Interacción entre las políticas, la ciencia abierta y la gobernanza de los datos de investigación en el contexto de la investigación científica y académica

Fuente: Elaboración propia

En este sentido, las políticas son un conjunto de actividades que se asocian con la toma de decisiones. Tratan de orientaciones o directrices que rigen la actuación de una persona o entidad en un asunto o campo determinado (RAE: 2023). Por lo tanto, las políticas pueden ser de diverso orden y naturaleza. En el contexto de la gobernanza de los datos de investigación, estas políticas permiten definir el manejo de los datos para su posterior accesibilidad a través del entorno digital.

De acuerdo con Solomonides (2019: 292), la gobernanza de los datos de investigación comprende los principios, políticas y estrategias adoptadas por las instituciones para proporcionar mecanismos operativos y de acceso a datos. Por lo tanto, las políticas para establecer la gobernanza de los datos conllevan el seguimiento de los principios que rigen a la recopilación, organización, transformación, el análisis, la recuperación y reutilización de los datos de investigación.

Con el propósito de ejercer una diferenciación entre las características de los conceptos de gobernanza, gobernanza de datos y gobernanza de datos de investigación, en la *Tabla 1* se ofrece una explicación de los componentes conceptuales de ambos elementos.

Concepto	Características
Gobernanza	<ul style="list-style-type: none"> • Es un concepto que se desprende de la doctrina política y sociológica. • Explica la interacción entre actores involucrados en la toma de decisiones de asuntos públicos. • En su aplicación, figuran las políticas como instrumentos para ejecutar y fundamentar la toma de decisiones.
Gobernanza de datos (GD)	<ul style="list-style-type: none"> • En su corpus, están presentes elementos como la calidad y confiabilidad de los datos. • Es un concepto que expone el papel de las políticas y los procesos que permiten garantizar el acceso a los datos. • Se relaciona con el término de gobierno de datos, debido a su aplicación en diversas instituciones de índole pública y privada.
Gobernanza de datos de investigación (GDI)	<ul style="list-style-type: none"> • Es un concepto que remite a un proceso intelectual para organizar y sistematizar datos con el propósito de hacerlos accesibles y reutilizables en el contexto digital. • Refleja la importancia de la reutilización de los datos para propiciar el desarrollo de nuevos conocimientos. • Su aplicación permite definir planes para la gestión de los datos de investigación disponibles en las instituciones académicas y científicas.

Tabla 1. Características de los conceptos de gobernanza, gobernanza de datos y gobernanza de datos de investigación

Elaboración propia. Fundamentada en las aportaciones de Cristina Zurbruggen (2011), T. Koltay (2016), y Rene Abraham, Johannes Schneider y Jan vom Brocke (2019)

De acuerdo con Abraham (2019: 426) “la gobernanza de los datos especifica un marco multifuncional para administrar datos como un activo organizacional estratégico. Al hacerlo, la gobernanza de los datos especifica los derechos de decisión y responsabilidades para la toma de decisiones de una organización sobre sus datos. Además, la gobernanza de datos formaliza las políticas, estándares y procedimientos aplicables a los datos supervisando su cumplimiento”.

Bajo esta premisa, la GD es un concepto genérico que, al momento de aplicarse en el contexto de la investigación científica y académica, remite a la presencia de los datos de investigación, en el sentido de su manejo con el propósito de fomentar su accesibilidad y reutilización. Por lo tanto, uno de los tópicos que la GDI aborda es la organización y la sistematización de los datos de investigación, pues estos procesos son claves para propiciar su accesibilidad y consulta en el contexto digital.

La GDI es un concepto holístico que no es exclusivo para el tratamiento de un solo tipo de datos de investigación. Pues la naturaleza de estos datos es ampliamente diversa y compleja, por lo que los procesos de organización y sistematización deben fundamentarse en el análisis de los atributos que presenten los datos de investigación.

Organización y sistematización de los datos de investigación

“Los datos de investigación se recopilan, observan, registran o crean para fines de análisis y producir resultados de investigación originales para un estudio” (Henderson, 2017: 2). El uso de metadatos para organizar a los datos de investigación es un tema que se ha planteado con anterioridad dentro de la literatura especializada. Si bien los metadatos son elementos descriptivos que permiten representar, recuperar y hacer accesibles los recursos y contenidos informativos disponibles en el entorno digital, cabe señalar que la concepción de la sistematización de los datos de investigación refleja un cambio sustancial en la manera de efectuar su organización.

Por ejemplo, Redkina (2019: 53) desarrolló un análisis de las políticas, directrices y lineamientos que son generados en diversos países cuyo propósito es dar mayor apertura a los datos de investigación mediante la aplicación de metadatos que forman parte de un plan para la gestión de datos de investigación. En este trabajo resalta la necesidad de contar con esquemas de metadatos específicos para organizar conjuntos de datos de investigación.

Por otra parte, Crystal-Ornelas *et al.* (2022: 1) llevaron a cabo un estudio de los datos que son generados en el contexto de las Ciencias de la Tierra y del Medio Ambiente, esto en el sentido de identificar la aplicación de los metadatos que utilizan para describir a los conjuntos de datos que son generados por

diversas comunidades científicas de ese campo disciplinario y hacerlos interoperables con los principios FAIR (Findability: *Localización*, Accessibility: *Accesibilidad*, Interoperability: *Interoperabilidad*, Reuse: *Reutilización*).

“Los principios FAIR hacen énfasis en mejorar la capacidad de las computadoras (es decir, la capacidad de los sistemas computacionales para encontrar, acceder, interoperar y reutilizar datos con una intervención humana mínima o nula) esto debido a que los humanos confían cada vez más en el soporte computacional para manejar los datos como resultado del aumento en el volumen, complejidad y velocidad de creación de los datos” (GoFAIR, 2023: párr. 1).

La capacidad de los metadatos para establecer un mecanismo de citación de datos otorga la posibilidad de contemplar los principios éticos relacionados con la certeza y calidad de los datos que son representados mediante los metadatos. Pues esto aumenta el nivel de confiabilidad y certeza al momento de respetar la propiedad intelectual y dar el debido reconocimiento a los creadores de los datos.

En este sentido, los datos, al momento de organizarse, “deben manejarse con cuidado durante su posesión y distribución. La propiedad y el control sobre los datos no son de naturaleza absoluta. Incluso el propietario principal, la fuente de los datos, sólo tiene derecho a utilizar los datos de forma ética. La propiedad de los datos se refiere tanto a la posesión como a la responsabilidad de la información” (Shukla et al., 2022: 69).

Desde esta perspectiva ética, la gobernanza de los datos de investigación puede ser útil para dotar de una mayor organización y sistematización a los datos; esto mediante la generación de políticas que regulen la sistematización de dichos datos.

Algunos ejemplos de estas regulaciones pueden observarse en proyectos como R3Data (<https://www.re3data.org/>), DataONE (<https://www.dataone.org/>) y Dryad (<https://datadryad.org/stash>), pues son proyectos que están basados en diversos lineamientos y políticas que, si bien son aplicables a diversas aristas de los datos de investigación, abordan diferentes temas respecto a su manejo y organización.

La sistematización de los datos de investigación pone de manifiesto una visión holística relacionada no sólo con la generación de repositorios de datos, sino también de instrumentos que regulen su apertura tomando en cuenta los factores técnicos, legales y económicos que forman parte del contexto académico y científico. En este sentido, la confiabilidad de un repositorio de datos debe enfocarse en:

- Adherirse a los estándares de metadatos y conservación, además de proporcionar la administración de los datos que almacena, por ejemplo, validación técnica, documentación, control de calidad, protección de la autenticidad y persistencia a largo plazo.

- Proporcionar servicios de datos, por ejemplo, interfaces de visualización, descarga de datos y procesamiento basado en servidor.
- Administrar los derechos de propiedad intelectual de los creadores de datos, la protección de los recursos de información sensible y la seguridad del sistema y su respectivo contenido. (Lin et al., 2020: 3)

En este sentido, la confiabilidad de la sistematización y organización de los datos dotará de una mayor certeza a los usuarios para poder utilizarlos, siendo este un principio relacionado con la gobernanza de los datos de investigación al momento de ser colocados en el entorno digital para su respectiva apertura, accesibilidad y re-utilidad.

METODOLOGÍA

La metodología empleada en este trabajo se desarrolló en dos niveles. En el primer nivel, se llevó a cabo un proceso de revisión de la literatura, apoyado en la hermenéutica y el análisis del discurso, con la intención de analizar recursos de información vinculados al objeto de estudio planteado en la investigación. De esta manera el análisis argumentativo permitió interpretar los textos referentes al objeto de estudio, en función de la idea principal planteada, sus antecedentes y el propósito de su comunicación.

Para esto, se analizaron artículos de investigación, capítulos de libro y estudios de caso. Estos recursos de información fueron obtenidos en diversas bases de datos, como Elsevier, Springerlink, IEEE y Emerald; en repositorios institucionales, como el repositorio institucional de la UNAM y el repositorio de la CEPAL; y, además, en sitios web educativos e institucionales especializados en el objeto de estudio planteado en la investigación (Research Data Alliance, Research Data Management @Harvard). La selección de estas fuentes responde a su especificidad temática relacionada con el manejo y la gobernanza de los datos de investigación, pues son fuentes que tratan de manera particular el estudio de los datos desde diversas aristas, por ejemplo, procedimental, teórica y tecnológica.

Dentro de estas fuentes, se implementaron estrategias de búsqueda relacionadas con los siguientes términos: *gobernanza de datos*, *gobernanza de datos de investigación*, *organización de datos* y *sistematización de datos* y sus equivalentes en inglés. En la *Tabla 2*, se expone este proceso con el número de registros obtenidos.

Términos de búsqueda	Springer-Link	Emerald Insight	Elsevier	IEEE Xplore	Repositorio UNAM	Repositorio CEPAL
<i>Gobernanza de datos</i>	82	16	287	0	244	2,882
<i>Data governance</i>	237,411	62,000	136,628	3,431	21	11,409
<i>Gobernanza de datos de investigación</i>	55	11	224	0	76	2,774
<i>Research data governance</i>	214,728	60,000	127,901	1,447	6	9,640
<i>Organización de datos</i>	721	111	15,670	0	462	17,518
<i>Data organization</i>	1,837,646	209,000	1,830,628	45,141	159	11,133
<i>Sistematización de datos</i>	47	2	672	0	39	3,885
<i>Data systematization</i>	19,329	1,000	30,750	172	9	4,853

Tabla 2. Número de registros obtenidos a partir de las estrategias de búsqueda relacionadas con el objeto de estudio de la investigación
Fuente: Elaboración propia

Cabe señalar que, aunque los términos empleados permitieron recuperar recursos que emplearan dichos términos dentro de su título, fue necesario revisar detalladamente cada recurso para identificar el enfoque disciplinario al que estaban dirigidos. Pues la noción conceptual de datos representa a múltiples variables de estudio, desde aquellas que versan sobre el rubro técnico y pragmático, hasta aquellas que permiten considerar sus nociones epistemológicas. En el caso de los sitios Research Data Alliance y Research Data Management @Harvard, no se realizaron búsquedas específicas de recursos, sino que se llevó a cabo un análisis de la documentación que contenían en sus páginas web que tratara los temas planteados en el objeto de estudio de esta investigación. De esta manera, la hermenéutica y el análisis del discurso permitió llevar a cabo la interpretación de los fenómenos particulares (gobernanza de datos de investigación) que acontecen en un contexto de tiempo y espacio definido (Cárcamo, 2005: 204).

Por lo tanto, se seleccionaron recursos publicados de 2015 a 2022, tomando en cuenta criterios que reflejasen una condicionante tecnológica y cronológica, pues el estudio de los datos se trata de un fenómeno que se encuentra en un constante dinamismo debido a la aparición de múltiples publicaciones que analizan su impacto

en diferentes contextos y disciplinas. Y, además, la mayoría de los recursos utilizados fueron publicados en idioma inglés.

La periodicidad de los años seleccionados responde a la constante actualización de los contenidos y los tópicos que se presenta en el objeto de estudio de los datos, pues en años recientes la literatura que versa sobre este tema ha crecido exponencialmente. Se optó por que los recursos seleccionados incluyeran, dentro de su espectro argumentativo, temas como *gobernanza de datos*, *gobernanza de datos de investigación*, *organización de datos* y *sistematización de datos*, para someterlos al análisis del discurso contenido dentro de su estructura argumentativa.

Los recursos de información publicados fuera del rango periódico mencionado fueron elegidos tomando en cuenta su grado de contribución y pertinencia respecto al análisis del objeto de estudio planteado en este trabajo; es decir, tomando en cuenta su relevancia para obtener una visión global del tema y que se relacionara con los tópicos de organización de la información y el conocimiento.

A su vez, tomando como sustento el método analítico-sintético, se seleccionó y descargó un conjunto de datos de investigación del directorio de repositorios de datos de investigación Dryad, disponible en <https://datadryad.org/stash>, con la finalidad de describirlo y representarlo mediante el uso de Datacite Metadata Schema, versión 4.4, esto con el propósito de ejemplificar las fases que intervienen en la organización y sistematización de los datos de investigación, mediante el uso de un esquema interoperable y estandarizado internacionalmente.

Los pasos a seguir para la tarea de organización y representación de estos datos se pueden consultar en la *Tabla 3*.

Fase	Descripción
1. <i>Selección del conjunto de datos de investigación</i>	Se seleccionaron conjuntos de datos con licencia abierta, que permitiera descargar y reutilizar a los datos de investigación o bien depositarlos en un repositorio respetando sus derechos de creación y propiedad intelectual.
2. <i>Análisis de los atributos del conjunto de datos</i>	Se identificó la naturaleza y temática de los datos descargados, mediante un análisis que permitiera obtener las propiedades representativas del conjunto.
3. <i>Descripción del conjunto de datos de investigación</i>	Tomando en cuenta los elementos descriptivos del esquema de metadatos, se describieron los atributos de los datos de investigación, siguiendo directrices señaladas en dicho esquema.
4. <i>Representación del conjunto de datos de investigación</i>	Mediante el diseño de un perfil descriptivo de metadatos, se representaron los datos de investigación, asignando puntos de acceso para su latente búsqueda y recuperación.

5. <i>Sistematización del conjunto de datos de investigación</i>	Las descripciones realizadas fueron almacenadas en un silo de datos para identificar las variables de la sistematización de los datos en un entorno digital.
--	--

Tabla 3. Metodología para la organización y sistematización de datos de investigación

Fuente: Elaboración propia

De esta manera, la elección de DataCite Metadata Schema, responde a su nivel de especificidad para describir y representar datos de investigación, al ser un esquema desarrollado ex profeso para este objetivo. Mediante el uso de este esquema pueden obtenerse metadatos para describir atributos de datos de investigación de diversa naturaleza, desde datos numéricos, textuales y alfa-numéricos.

Sin embargo, el esquema no puede concebirse como un modelo genérico para la descripción de datos de investigación, pues su sentido holístico permite implementarlo en diversos contextos disciplinarios acordes con la propia naturaleza de los datos; por esta razón, se sugiere primeramente analizar y conocer los atributos que caracterizan a los datos que serán organizados y sistematizados.

La metodología para el manejo de datos de investigación, desde la perspectiva de la GDI, aporta elementos para la elaboración de planes para la gestión de dichos datos, lo cual puede contribuir a visibilizar en mayor medida el trabajo de los investigadores. De acuerdo con Palsdottir (2021: 323), entre los beneficios de la gestión de datos está que facilita el intercambio de datos, lo que permite que otros investigadores verifiquen los resultados y reutilicen los datos.

ANÁLISIS DE RESULTADOS

Al momento de descargar el conjunto de datos de investigación, fue preciso analizarlos y tener la posibilidad de consultarlos mediante el uso de *software* de visualización. Mediante el examen de las propiedades del conjunto, fue posible identificar los atributos de los datos y representarlos en un perfil de metadatos diseñado con DataCite Metadata Schema (véase la *Tabla 4*). De esta manera, fue posible aplicar una estructura de metadatos susceptible de fomentar la organización de los datos de investigación que fueron descritos como parte de este ejercicio.

El esquema de metadatos de DataCite, en su versión núcleo 4.4, está conformado por 20 elementos de metadatos, divididos en tres diferentes niveles de uso: Obligatorio, Recomendado y Opcional.

Mediante el uso del esquema base de metadatos fue necesario establecer un perfil de metadatos para describir y representar al conjunto de datos denominado “*Data from: Exploratory and confirmatory research in the open science era*”, tal y como se observa en la *Tabla 4*.

ID	Elemento	Descripción
1	Identificador	https://doi.org/10.5061/dryad.z612jm686
2.1	Nombre del creador	Nilsen, Erlend B.
2.4. b	SchemeURI	https://orcid.org/0000-0002-5119-8331
2.5	Afiliación	Norwegian Institute for Nature Research
2.1	Nombre del creador	Bowler, Diana.
2.5	Afiliación	German Center for Integrative Biodiversity Research
2.1	Nombre del creador	Linnell, John.
2.5	Afiliación	Norwegian Institute for Nature Research
3	Título	Data from: Exploratory and confirmatory research in the open science era
4	Editor	Dryad
5	Año de publicación	2020
6	Tema	Ciencia abierta
6	Tema	Revisión de la literatura
6	Tema	Investigación ecológica
8	Fecha	2020-02-03
8.a	Tipo de fecha	Publicación
9	Idioma	Inglés
10	Tipo de recurso	Conjunto de datos
12	Identificador relacionado	https://doi.org/10.1111/1365-2664.13571
12. a	Tipo de identificador relacionado	DOI
12. b	Tipo de relación	Es derivado de
13	Tamaño	70.3 kb
14	Formato	CSV
16	Derechos	Este trabajo tiene una licencia CC0 1.0 Universal (CC0 1.0) Public Domain Dedication .
16. a	URI de derechos	https://creativecommons.org/publicdomain/zero/1.0/

17	Descripción	Applied ecological research is increasingly inspired by the Open Science movement. However, new challenges about how we define our science when biodiversity data is being shared and re-used are not solved. Among these challenges is the risk associated with blurring the distinction between research that mainly seeks to explore patterns with no a-priori articulated hypotheses (exploratory research), and research that explicitly tests a-priori formulated hypotheses (confirmatory research).
17 a.	Tipo de descripción	Resumen
19	Financiamiento	Norges Forskningsråd, Award: 251112

Tabla 4. Descripción y representación de un conjunto de datos de investigación mediante un perfil de metadatos diseñado con DataCite Metadata Schema
Fuente: Elaboración propia, con datos de DataCite Metada Schema, versión 4.4. 2021.

El registro de datos en cada uno de los elementos descriptivos del perfil fue llevado a cabo mediante el principio de “registro de observación”, es decir, de la fuente original del recurso, se tomaron los datos que representaban los atributos del *dataset* que fue examinado en el repositorio Dryad a través de su descarga. El perfil de metadatos emplea un nivel de ocurrencia que está relacionado con la repetibilidad de un determinado metadato. En el caso de los elementos 2.1 y 2.5, esta variable puede verse reflejada al momento de describir el nombre de los creadores y su respectiva afiliación.

Cabe señalar que el diseño de cada perfil de metadatos debe desarrollarse tomando en cuenta el análisis de los atributos que representan a cada recurso de información, en este caso a cada conjunto de datos de investigación que sea depositado y almacenado en un sistema. A su vez, el elemento 2.4. *b* refiere al Schema URI y permite apreciar las vinculaciones que pueden llevarse a cabo para identificar a los autores con sus respectivos identificadores digitales. En este caso el identificador empleado por el esquema particular de ORCID.

El nivel de exhaustividad al momento de llevar a cabo la descripción y representación del conjunto de datos de investigación permite identificar las propiedades que forman parte de la naturaleza dinámica del recurso, es decir, algunos de los datos registrados en el perfil son susceptibles de actualizarse con frecuencia dentro del propio sistema Dryad.

La gobernanza de los datos de investigación es un concepto que permite remitirse al uso de estándares de metadatos que faciliten la organización de los datos, mediante el análisis de sus atributos. Dicha gobernanza en su nivel más abstracto hace posible la interacción de diversos elementos normativos e informativos. Por ejemplo, la generación de políticas para regular y uniformar la representación de los datos es un escenario global.

A nivel de sistematización, la GDI supone un cambio en cuanto a la generación de sistemas para buscar y recuperar datos, por ejemplo, la generación de repositorios que, en un escenario global, están comenzando a desarrollarse en diversas instituciones académicas y de investigación a nivel mundial.

Si bien el uso de repositorios institucionales para fomentar el acceso abierto a las publicaciones científicas y académicas ha cobrado una notable madurez, el fenómeno de la generación vertiginosa de datos y su uso en la investigación está cobrando doble relevancia al momento de considerar la GDI como un mecanismo para propiciar la accesibilidad y la reutilización de los datos de investigación. Por lo tanto, los resultados obtenidos mediante el ejercicio planteado en esta investigación permiten obtener un acercamiento en cuanto a la descripción y representación de datos de investigación, mediante el uso de un esquema interoperable.

DISCUSIÓN

En los resultados obtenidos, puede percibirse la aplicación de elementos de la gobernanza de los datos de investigación, los cuales son susceptibles de implementarse al momento de utilizar un esquema de metadatos formal para la descripción de datos de investigación. En este sentido, la GDI refleja un factor de especificidad mayor en cuanto al uso de principios jerárquicos para representar datos específicos. Por lo tanto, la GDI permite definir roles, procesos y puntos de acción para describir los datos de investigación.

A su vez, la GDI es un concepto que, al momento de implementarse pragmáticamente en la organización y sistematización de los datos, hace posible concebir la interrelación que debe existir entre las políticas y las directrices para describir y representar a los datos de una determinada institución. Este argumento fue adoptado como guía para el desarrollo de esta investigación, pues este tipo de gobernanza tiene un carácter holístico que permite contemplar la accesibilidad a los datos de investigación en un marco común, pero no totalitario ni genérico.

En el contexto de la ciencia abierta, la gobernanza de los datos de investigación tiene la particularidad de fomentar la interoperabilidad normativa y procedimental para construir un ecosistema de datos de investigación accesible para la sociedad y las comunidades que así lo requieran.

La propuesta que se presenta para la descripción y representación de los datos de investigación no es de carácter exhaustivo, pues el diseño del perfil de metadatos desarrollado con DataCite Metadata Schema tiene el propósito de ejemplificar los alcances para la organización y sistematización de dichos datos, ya que permite establecer un mecanismo interoperable para describir a los datos de investigación y su aplicación en un repositorio de datos, lo cual permite, a su vez,

dotar de mayor visibilidad a los datos generados en investigaciones científicas y académicas, además de propiciar su reutilización a través de un aparato de citación que propicie el respeto a la propiedad intelectual de dichos datos.

A su vez, la concepción de la gobernanza de los datos de investigación aún presenta limitaciones relacionadas con su propia epistemología, pues no queda claro su nivel de dominio dentro de las organizaciones académicas y científicas, y su importancia para propiciar una mayor comprensión de los problemas y fenómenos que caracterizan a la organización y la sistematización de los datos de investigación.

Al respecto, Mahanti (2021: 19) manifiesta que “existen factores, tanto internos como externos a una organización que pueden afectar el valor de los datos. Los datos pueden transformarse de un activo a un gran pasivo si no son aptos para su propósito, es decir, su calidad no está a la altura, por ejemplo, datos incorrectos o desactualizados, registros faltantes o datos que no son oportunos”.

Por lo tanto, la gobernanza de los datos de investigación se concentra en el manejo de los datos desde su perspectiva legal hasta la operativa, lo cual lleva a considerar su función dentro de los procesos de su organización y sistematización dentro de las instituciones científicas y académicas.

CONSIDERACIONES FINALES

Como parte de la investigación realizada, se considera que la GDI es un concepto caracterizado por dos componentes, por un lado, su visión política, que refleja la obligación de transparentar el uso de los datos que emanan de la actividad pública y privada, así como de las instituciones de gobierno y, por el otro, su postura relativa al manejo y gestión de los datos en dichas instituciones. Es dentro de este rubro en donde los procesos de la organización y sistematización de los datos de investigación tienen una interacción conceptual.

Por otra parte, la gobernanza de los datos de investigación busca establecer estándares internos y externos en las instituciones de investigación científica y académica para dotar de una mayor calidad y confiabilidad a los datos con fines de su accesibilidad y reutilización. Esto conlleva el establecimiento de políticas para designar una infraestructura capaz de asignar roles y responsabilidades para manejar, organizar y sistematizar dichos datos.

A su vez, la organización y sistematización de los datos de investigación son dos procesos que interactúan en la gobernanza de los datos, pues forma parte del manejo y gestión que los datos deben tener al momento de pensar en su acceso y reutilización. En este sentido, un perfil de metadatos diseñado ex profeso para este tipo de datos puede contribuir a establecer un mecanismo interoperable para

describir y representar a los datos de investigación en un sistema disponible en el contexto digital.

De esta manera, la gobernanza de los datos de investigación es un tema con diversas aristas. En este trabajo se ha presentado aquella que tiene lugar en la organización y sistematización de los datos, mediante un esquema desarrollado ex profeso para llevar a cabo la descripción y la representación de los datos a través de un enfoque sistematizado, lo cual permite advertir que los datos de investigación no pueden ser tratados de una manera generalizadora, sino llevando a cabo el análisis de sus atributos para poder representarlos de acuerdo a los requerimientos de un determinado contexto disciplinario.

REFERENCIAS

- Abraham, Rene, Johannes Schneider, y Jan vom Brocke. 2019. "Data Governance: A Conceptual Framework, Structured Review, and Research Agenda". *International Journal of Information Management* (49): 424-438.
<https://doi.org/10.1016/j.ijinfomgt.2019.07.008>.
- AmeliCA. 2019. "Ciencia Abierta". Acceso 1 de febrero de 2023.
<http://www.amelica.org/proyectos/ciencia.html>
- Cárcamo Vásquez, H. 2005. "Hermenéutica y análisis cualitativo". *Cinta de Moebio: Revista de Epistemología de Ciencias Sociales* 23 (1): 204-216.
<https://www.moebio.uchile.cl/23/carcamo.html>
- Crystal-Ornelas, Robert, Charuleka Varadharajan, Dylan O’Ryan, Kathleen Beilsmith, Benjamin Bond-Lamberty, Kristin Boye, Madison Burrus, et al. 2022. "Enabling FAIR Data in Earth and Environmental Science with Community-Centric (Meta) Data Reporting Formats". *Scientific Data* 9 (1): 700.
<https://doi.org/10.1038/s41597-022-01606-w>
- DataCite Metadata Working Group. 2021. "DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs v4.4". Application/pdf.
<https://doi.org/10.14454/3W3Z-SA82>
- Fleckenstein, Mike, y Lorraine Fellows. 2018. "Data Governance". En *Modern Data Strategy*, editado por Mike Fleckenstein y Lorraine Fellows, 63-76. Cham: Springer International Publishing.
https://doi.org/10.1007/978-3-319-68993-7_8
- Go FAIR. 2023. "FAIR Principles". Acceso 2 de febrero de 2023.
<https://www.go-fair.org/fair-principles/>
- Henderson, Margaret E. 2017. "Data management: a practical guide for librarians". *Practical guides for librarians*, no. 28. Lanham, Maryland: Rowman & Littlefield.
- Jansen, M., Paul Brous, Elsa Estévez, Luis Barbosa, y Tomasz Janowski. 2020. "Data Governance: Organizing Data for Trustworthy Artificial Intelligence". *Government Information Quality* 37,
<https://doi.org/10.1016/j.giq.2020.101493>

- Karmanovskiy, Nikolay, Dmitry Mouromtsev, Mikhail Navrotskiy, Dmitry Pavlov, e Irina Radchenko. 2016. "A Case Study of Open Science Concept: Linked Open Data in University". En *Digital Transformation and Global Society*, editado por Andrei V. Chugunov, Radomir Bolgov, Yuri Kabanov, George Kampis y Maria Wimmer, 400-403. Communications in Computer and Information Science. Cham: Springer International Publishing.
https://doi.org/10.1007/978-3-319-49700-6_39
- Koltay, T. 2016. "Data governance, data literacy and the management of data quality". *IFLA Journal* 42 (4).
<https://journals.sagepub.com/doi/full/10.1177/0340035216672238>
- Kuiler, Erik W. 2022. "Data Governance". En *Encyclopedia of Big Data*, editado por Laurie A. Schintler, y Connie L. McNeely, 286-290. Cham: Springer International Publishing.
https://doi.org/10.1007/978-3-319-32010-6_306
- Lin, Dawei, Jonathan Crabtree, Ingrid Dillo, Robert R. Downs, Rorie Edmunds, David Giarretta, Marisa De Giusti, et al. 2020. "The TRUST Principles for Digital Repositories". *Scientific Data* 7 (1): 144.
<https://doi.org/10.1038/s41597-020-0486-7>
- Mahanti, Rupa. 2021. "Data and Its Governance". En *Data Governance and Data Management: Contextualizing Data Governance Drivers, Technologies, and Tools*, editado por Rupa Mahanti, 5-82. Singapore: Springer.
https://doi.org/10.1007/978-981-16-3583-0_2
- OECD. 2022. "Going Digital to Advance Data Governance for Growth and Well-Being".
<https://doi.org/10.1787/e3d783b0-en>
- Palsdottir, Agustá. 2021. "Data literacy and management of research data a prerequisite for the sharing of research data". *Aslib Journal of Information Management* 73 (2): 322-41.
<https://doi.org/10.1108/AJIM-04-2020-0110>
- RAE. 2023. "Política". En *Diccionario de la lengua española*. Acceso 1 de febrero de 2023.
<https://dle.rae.es/política>
- Redkina, N. S. 2019. "Current Trends in Research Data Management". *Scientific and Technical Information Processing* 46 (2): 53-58.
<https://doi.org/10.3103/S0147688219020035>
- Shukla, Samiksha, Jossy P. George, Kapil Tiwari, y Joseph Varghese Kureethara. 2022. "Data Ethics". En *Data Ethics and Challenges*, editado por Samiksha Shukla, Jossy P. George, Kapil Tiwari y Joseph Varghese Kureethara, 61-72. Springer Briefs in Applied Sciences and Technology. Singapore: Springer.
https://doi.org/10.1007/978-981-19-0752-4_4
- Solomonides, A. 2019. "Research Data Governance, Roles, and Infrastructure". En *Clinical Research Informatics*, editado por Rachel L. Richesson y James E. Andrews, 291-310. Health Informatics. Cham: Springer International Publishing.
https://doi.org/10.1007/978-3-319-98779-8_14
- Wang, Chen-Shu, Shiang-Lin Lin, Tung-Hsiang Chou, y Bo-Yi Li. 2019. "An Integrated Data Analytics Process to Optimize Data Governance of Non-Profit Organization". *Computers in Human Behavior* 101 (diciembre): 495-505.
<https://doi.org/10.1016/j.chb.2018.10.015>
- Weber, Kristin, Boris Otto, y Hubert Österle. 2009. "One Size Does Not Fit All: A Contingency Approach to Data Governance". SSRN Scholarly Paper: Rochester, NY.
<https://papers.ssrn.com/abstract=1728505>

Zurbriggen, Cristina. 2011. “Gobernanza: una mirada desde América Latina”. *Perfiles latinoamericanos* 19 (38): 39-64.

Para citar este texto:

Ávila-Barrientos, Eder. 2023. “La gobernanza de los datos de investigación en el contexto de su organización y sistematización”. *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 37 (96): 45-64.

<http://dx.doi.org/10.22201/iibi.24488321xe.2023.96.58763>