
ESTUDIOS / RESEARCH STUDIES

Estructura propuesta del artículo de datos como publicación científica

Sandra M. Roa-Martínez*., Silvana A. B. Vidotti*, Ricardo C. Santana*

*Universidade Estadual Paulista Júlio de Mesquita Filho (Brasil) - Grupo de Pesquisa Novas Tecnologias em Informação

Correo-e: smroa77@gmail.com | ORCID iD: <http://orcid.org/0000-0002-2271-6101>

ORCID iD: <http://orcid.org/0000-0002-4216-0374> | ORCID iD: <http://orcid.org/0000-0003-1387-4519>

** Universidad del Cauca (Colombia) – Grupo de Investigación en Inteligencia Computacional

Correo-e: smroa@unicauca.edu.co

Recibido: 29-02-2016; 2ª versión: 26-06-2016; Aceptado: 30-06-2016.

Cómo citar este artículo/Citation: Roa-Martínez, S. M.; Vidotti, S. A. B.; Santana, R. C. (2017). Estructura propuesta del artículo de datos como publicación científica. *Revista Española de Documentación Científica*, 40(1): e167. doi: <http://dx.doi.org/10.3989/redc.2017.1.1375>

Resumen: Este trabajo presenta una revisión de las principales motivaciones y caminos en la publicación de los conjuntos de datos, siendo estos conjuntos de datos generados y manipulados durante los procesos de investigación. Se considera el Artículo de datos como medio de publicación científica con igual reconocimiento, aceptación y rigor científico que un artículo convencional o tradicional de investigación actual. Por cuanto se propone una estructura común definida con elementos principalmente derivados de los metadatos del conjunto de datos de investigación. Esto permitirá a los creadores, editores, consumidores y pares expertos evaluadores el reconocimiento, compartición, evaluación y reutilización de datos. Se facilita así la reproducibilidad de la información, validación de los resultados divulgados y rápida generación de nuevas investigaciones.

Palabras clave: Artículo de datos; publicación de datos de investigación; compartición de datos.

Proposed structure of a data paper structure as scientific publication

Abstract: This paper presents a review of the main motivations and paths for publishing datasets that are generated and managed during the research process. The Data Paper is considered as a form of scientific publication with the same recognition, acceptance and scientific rigor as conventional research articles. Therefore we propose a common structure defined by elements based mainly on dataset metadata. This will enable creators, publishers, consumers and expert peer reviewers to recognise, share, evaluate and facilitate data reuse. Doing so will facilitate information reproducibility, validation of results, and rapid new research generation.

Keywords: Data paper; published data; research data; data sharing.

Copyright: © 2017 CSIC. Este es un artículo de acceso abierto distribuido bajo los términos de la licencia *Creative Commons Attribution (CC BY)* España 3.0.

1. INTRODUCCION

Dada la importancia de los datos de investigación para el desarrollo de la ciencia, ya que a partir de ellos los investigadores obtienen resultados tras su análisis y utilización, es relevante que los datos sean compartidos con otros investigadores a través de las Tecnologías de Información y Comunicación (TIC). Mediante la reutilización de dichos datos, podrían validarse los resultados y/o alcanzarse nuevas conclusiones derivadas de otros intereses sobre el mismo conjunto de datos pero analizado desde otra perspectiva. Sin embargo, para que estos datos sean útiles deben no solo compartirse sino publicarse.

Actualmente, la mayoría de los artículos científicos publicados presenta los resultados de un proceso de investigación y cuenta con una estructura similar en la que ciertos aspectos del formato, pero no de la estructura, se modifican según las pautas dadas a los autores por el periódico o revista donde será publicado. Además, los artículos se reconocen por la rigurosidad implícita del proceso de investigación y evaluación o revisión por pares expertos. La publicación de los datos que llevaron a estos resultados también sigue un proceso sistemático. Este proceso, puede ser reconocido, al igual que los resultados, mediante su publicación a través de un Artículo de Datos. Sin embargo, los Artículos de Datos no cuentan con una estructura común; han sido en ocasiones una adaptación del artículo convencional y en otras, un extenso o limitado conjunto de elementos que describen al conjunto de datos. Por otra parte, la falta de reconocimiento al esfuerzo del investigador en su generación se convierte en un factor que desmotiva a la hora de pensar en su publicación. El Artículo de datos puede estar sujeto a la revisión por pares expertos y contar con una estructura y formatos específicos para su publicación.

Por todo lo anterior, el objetivo principal de este trabajo es presentar una estructura común para el Artículo de Datos de investigación. Se presenta este tipo de artículo como medio de publicación a partir del análisis de un conjunto de editores que agrupan un número significativo de revistas de datos o *data journals* que actualmente publican este tipo de artículos. Asimismo, se presenta una revisión de guías, plantillas o formatos utilizados por comunidades familiarizadas con este tipo de publicaciones y con la documentación de los datos de investigación.

La metodología utilizada en este trabajo fue descriptiva, a partir de la compilación y sistematización de información cualitativa y cuantitativa de las directrices, guías y demás proporcionadas por las páginas web de los editores y revistas que fueron objeto de estudio de este trabajo. Además, se tuvo un enfoque prospectivo, ya que este trabajo sugiere una estructura que busca consolidarse a partir de los múltiples formatos existentes y otros docu-

mentos y guías que forman parte del estado de la práctica en cuanto a Artículos de Datos se refiere. La recolección de la información fue realizada visitando páginas web y documentos publicados.

En este documento, se abordan los principales conceptos relacionados dentro de los cuales se enmarcó el trabajo propuesto. En primer lugar, se delimitó lo que actualmente se consideran datos de investigación, presentándose la compartición de datos y otros aspectos relacionados a la publicación de datos. Posteriormente, se describe el Artículo de Datos y se propone una estructura de los elementos que deberá tener un Artículo de Datos de investigación a partir principalmente de los metadatos del conjunto de datos. Se finaliza con las conclusiones derivadas del trabajo realizado.

2. DATOS DE INVESTIGACIÓN

Si bien el concepto de dato puede atribuirse a una representación, una información, un suceso y un amplio conjunto de definiciones, es relevante delimitarlo e ir más allá, concentrándonos en el significado de los datos de investigación y, con ello, facilitar la discusión acerca de la compartición y publicación de los mismos.

A partir de la definición «Los datos son hechos, números, letras y símbolos que escriben un objeto, idea, condición, situación u otros factores» (CSPASTDPI y otros, 1999), y ampliándola según el interés de este artículo, los datos se consideran también como «Cualquier información que se puede almacenar en forma digital, incluyendo texto, números, imágenes, vídeo o películas, audio, software, algoritmos, ecuaciones, animaciones, modelos, simulaciones, etc., estos datos pueden ser generados por diferentes medios, entre ellos la observación, cálculo o experimento» (NSB y NSF, 2005). Con esto, aunque delimitado, es evidente la complejidad de todo aquello que puede ser considerado un dato, pero falta aún especificar cómo caracterizar los datos de investigación o científicos, los cuales son de principal interés para la comunidad de investigadores.

También existen amplios significados y terminología asociados a datos de investigación o científicos. Para facilitar la lectura, será adoptado en este trabajo el término datos de investigación abarcados en la siguiente definición, construida a partir de instituciones como National Institutes of Health (NIH, 2003) y Organisation for Economic Co-Operation and Development (OECD, 2007): «[...] todo aquel material que ha sido registrado durante la investigación, reconocido por la comunidad científica y que sirve para certificar los resultados de la investigación que se realiza [...] que debe provenir de una fuente única», sin tener en cuenta otros elementos

como las notas de laboratorio, los análisis previos, bocetos, informes o conversaciones informales con otros colegas (Torres-Salinas y otros, 2012).

Es importante tener en cuenta que los datos de investigación ni pueden ser caracterizados ni enmarcados de igual forma para las diferentes disciplinas o áreas de estudio de los investigadores. Sin embargo, para este trabajo, los datos de investigación generados por cualquier disciplina o investigador tendrán igual importancia y podrán ser susceptibles de ser publicados y compartidos por sus creadores con la comunidad interesada conllevando un mayor avance de la ciencia.

3. COMPARTICIÓN DE DATOS

La compartición de datos, también conocida como *Data Sharing*, ha sido definida por diversos autores como:

- «La acción de compartir con otros colegas los ficheros de datos (o *raw data*) generados durante el curso de una investigación» (Torres-Salinas, 2009).
- «Publicación de los datos de investigación para su uso por otros» (Borgman, 2012).

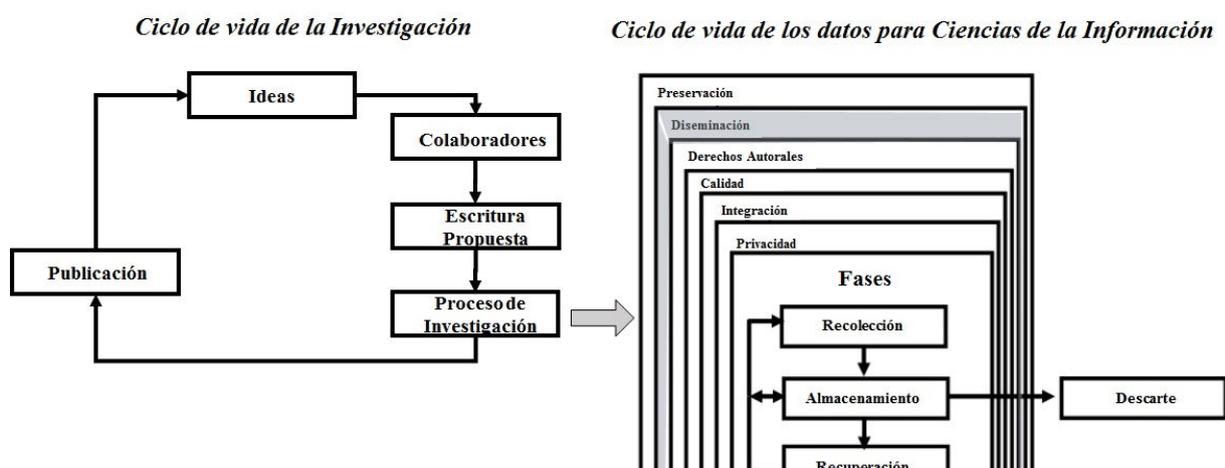
La compartición de datos puede ser entonces descrita como la acción de hacer disponibles los datos de investigación de diversas formas, desde un intercambio privado o directo por solicitud del interesado con el creador, o a partir del almacenamiento (e.j. un repositorio público de datos) para que sean usados con diferentes fines por otros investigadores.

Por otra parte, dada la importancia de los datos en la investigación, considerados como «la piedra angular de la Ciencia» (Wallis y otros, 2013), su ciclo de vida no puede ser considerado independiente de las etapas de la investigación.

El Ciclo de vida de la investigación de Tenopir y otros (2011) inicia con las ideas, pasa por la escritura de la propuesta y llega al proceso de investigación como una de sus etapas, dentro de la cual los datos deben surtir su propio ciclo de vida, por cuanto fue tomado como referencia para estos el ciclo de vida de los datos para Ciencias de la Información propuesto por Sant'Ana (2013). Este ciclo, a partir del análisis y revisión de modelos diversos, presenta como uno de sus objetivos la diseminación, la cual está correlacionada con las actividades de distribución y compartición de los datos. Asimismo, su fase de recuperación se asocia con el acceso, uso, reuso y análisis de los datos. Principalmente, este objetivo y esta fase mencionados son objeto de interés de este trabajo que busca garantizar que los datos sean compartidos e igualmente reusados por las comunidades académicas. En la figura 1, puede observarse una integración de los ciclos de vida de la investigación y de los datos para las Ciencias de la Información descritos.

Si bien ha sido ampliamente discutido en la literatura el tema de compartición de datos en los últimos tiempos; a partir de la información presentada en algunas de las referencias revisadas (Torres-Salinas y otros, 2012; Torres-Salinas, 2009; Tenopir y otros, 2011; Borgman, 2012; OECD, 2007; Van den Eynden y Bishop, 2014; Wallis y Borgman, 2013), se encuentra que son amplios e importantes los be-

Figura 1. Ciclos de vida de la investigación y de los datos para Ciencias de la Información



Fuente: Tenopir y otros, 2011; Sant'Ana, 2013, adaptación y traducción propia.

neficios de la compartición de datos de investigación, ya que, entre otros aspectos, se maximiza su potencial, se disminuyen costos asociados a los recursos requeridos y se evita duplicidad de investigaciones. Por otra parte, las desventajas encontradas pueden considerarse oportunidades de mejora a la espera de propuestas de soluciones. Estas propuestas pueden ser abordadas, en su mayoría, desde las Ciencias de la Información y están relacionadas con el acceso, los metadatos, la disponibilidad y la normalización de los procesos de almacenamiento, gestión y preservación de los datos de investigación. El objetivo primordial de la compartición de datos es disminuir el esfuerzo y el tiempo de los investigadores, entre otros recursos, a cambio de incrementar principalmente su reconocimiento.

4. PUBLICACIÓN DE DATOS

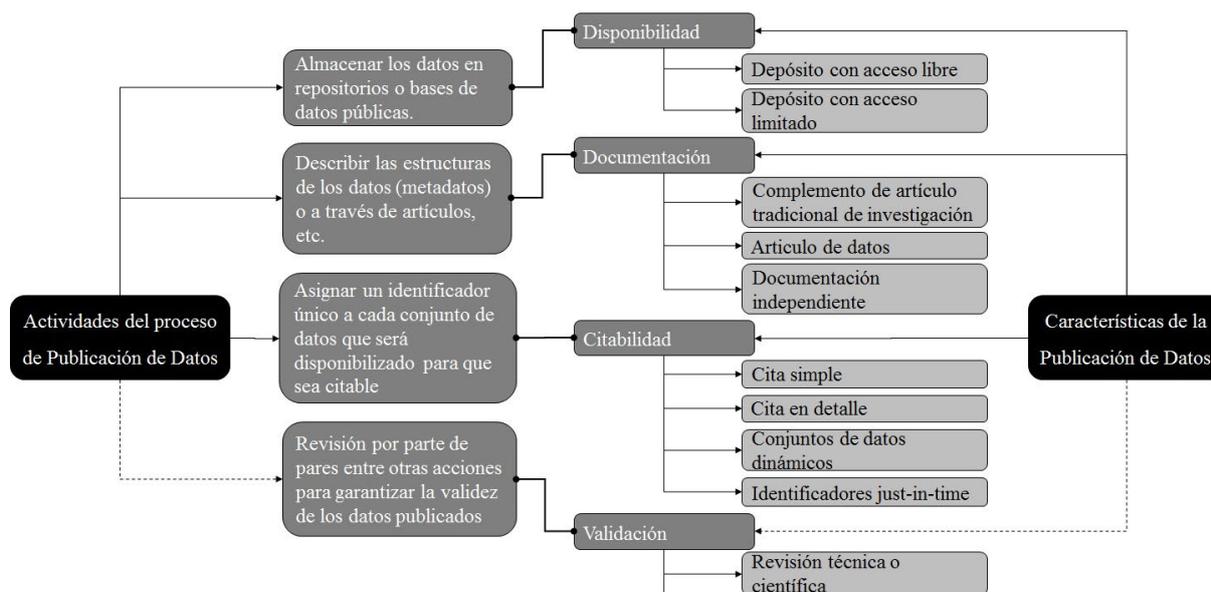
Al igual que los conceptos analizados previamente, la Publicación de Datos no es ajena a una multiplicidad de definiciones y puede llegar a ser entendida como Compartición de Datos, Datos Abiertos u *Open Data* y Liberación de Datos o *Data Release*. Por cuanto, publicación de datos se diferencia de Publicación de datos (Callaghan y otros, 2012; Lawrence y otros, 2011), no solo por una letra mayúscula o minúscula, sino porque la primera se refiere a cualquier tipo de dato que tiene como principales características el ser *compartido, disponible y distribuido*, mientras que las características que deben cumplir los datos de investigación para su Publicación según

Kratz y Strasser (2014) son: (1) los datos publicados deben ser disponibles al público en cualquier momento, (2) los datos publicados deben estar debidamente documentados para que sean fácilmente reproducibles o reutilizados y (3) una publicación de datos puede ser formalmente citada como los libros o artículos de una revista. Y otra propiedad, que hoy por hoy se encuentra abierta y en discusión, está relacionada con (4) la validación del conjunto de datos para que sean publicados.

Las actividades del proceso de Publicación del conjunto de datos de investigación, también mencionadas por Kratz y Strasser (2014), se encuentran estrechamente relacionadas con cada una de las características presentadas de los mismos autores: (1) almacenar los datos en repositorios o bases de datos públicas, (2) describir las estructuras de los datos (metadatos) a través de artículos, etc., (3) asignar un identificador único a cada conjunto de datos que será disponibilizado para que sea citable y (4) revisar por parte de pares, entre otras acciones, para garantizar la validez de los datos publicados. Las características y actividades asociadas al proceso de publicación de datos están actualmente reconocidas por la comunidad de creadores, editores y consumidores de datos de investigación publicados.

En la figura 2, han sido incorporadas las actividades del proceso de Publicación de datos relacionadas con las características que debe cumplir el conjunto de datos para su Publicación.

Figura 2. Proceso y características de la Publicación de Datos



Fuente: Kratz y Strasser, 2014, adaptación propia.

Adicionalmente, en la misma figura 2, se muestran mecanismos que representan a cada par actividad-característica. Un ejemplo es el Artículo de datos que permite describir las estructuras de los datos (actividad del proceso de Publicación) y forma parte de la Documentación (característica de la Publicación). Existen otros mecanismos en este mismo par (actividad-característica) como el complemento de un artículo tradicional u otro tipo de documentación independiente, así como otras actividades como almacenar los datos en repositorios o bases de datos públicas. Este almacenamiento garantiza la disponibilidad y puede ser realizado en depósitos con acceso libre o con acceso limitado. La asignación de un identificador único a cada conjunto de datos que será disponibilizado permite su citabilidad y será realizada por medio de citas simples, en detalle, datos dinámicos o identificador *just-in-time*. Finalmente, la validez de los datos publicados puede ser técnica o científica y realizada por pares expertos o independientes.

5. ESTRUCTURA PROPUESTA DEL ARTÍCULO DE DATOS

Esta sección inicia con una definición del Artículo de Datos, también conocido como *datapaper*, *data descriptor* o *article of data*. Se considera que «un Artículo de Datos describe una colección de datos, un proceso, un software, formatos de archivo, etc., sin el requisito de un análisis novedoso o de conclusiones innovadoras. Describe cuándo, cómo y por qué se recogieron los datos y en qué consiste el producto» (Whyte y otros, 2013; García-García y otros, 2015). Esto nos permite diferenciarlos notablemente de los artículos científicos convencionales, de revisión y notas breves. Es decir, un Artículo de datos sería un nuevo tipo de comunicación científica (Newman y Corke, 2009). El Artículo de Datos no puede ser considerado como material suplementario, ya que el alcance de este tipo de material no se extiende a aspectos relacionados con la gestión de datos en general, sino a la inclusión de los datos que son publicados como materiales suplementarios (NISO, 2013). Es decir, los procesos de recolección de datos, formatos y otros que forman parte de la gestión de datos sí serán considerados entre los elementos propuestos para la estructura del Artículo de Datos de este trabajo.

Consideramos el Artículo de Datos como un nuevo tipo de comunicación científica que forma parte de las clases de documentación producto del proceso de Publicación de datos. Este tipo de artículo describe un conjunto de datos de investigación accesibles por otros investigadores junto con información propia, es decir, con metadatos relacionados al proceso del que surgieron los datos desde

los métodos y consideraciones de su recolección, formatos, condiciones específicas y autores entre otros. La necesidad de la existencia de un Artículo de Datos está principalmente motivada por el potencial que tienen los datos. Además, permitirá validar los resultados obtenidos a partir de los datos mostrados, establecer nuevas hipótesis a partir de ellos, y dar reconocimiento a sus autores. Es decir, contribuirá a la solución de varias de las desventajas mencionadas con anterioridad y otros aspectos relacionados con la compartición.

Son múltiples los acercamientos existentes a una propuesta de estructura de un Artículo de Datos, pero la gran parte de éstas han sido adaptadas de los artículos tradicionales donde cada periódico o revista determina una plantilla o formato. En algunas ocasiones, este formato puede estar incompleto y, en otras, ser demasiado extenso. Por cuanto, este trabajo presenta una estructura donde aquellos elementos que no se incluyen serán opcionales, mientras que los descritos deberán ser necesarios. Con ello, se busca la flexibilidad dentro de un marco común que facilitará la validación y reutilización de los datos de investigación generados y publicados en un Artículo de Datos junto a otros beneficios que, a mediano plazo, podrán ser analizados (reconocimiento de los creadores, nuevas dinámicas de investigación a partir de la reutilización de los datos, etc.).

El principal aporte de este trabajo es proporcionar una estructura común del Artículo de datos a partir de la revisión de un grupo de revistas que en la actualidad publican este tipo de artículos. Estas revistas son conocidas como Revistas de Datos o *Data Journals* y han sido previamente analizadas con otros fines por Candela y otros (2015). En este trabajo, se pretende ampliar los resultados presentados allí, complementarlos y, finalmente, especificar el conjunto de elementos que compondrán una estructura común que deberían seguir los Artículos de Datos.

Metodológicamente, este análisis se inició con la revisión de 14 editores que agrupan 115 revistas de datos de diferentes disciplinas y áreas (Medicina, Química, Biología, Biodiversidad, etc.). Fue visitado el sitio web de cada editor y/o el de las revistas específicas y se ubicaron las informaciones de publicación proporcionadas por los mismos. El 64% proporcionaba formatos o directrices específicas (9 de 14 editores). Posteriormente, fue construido un conjunto de 36 elementos (secciones o informaciones) encontrados en sus guías para autores, directrices y plantillas con los cuales se construyó la Tabla I. En ella, se presentan los elementos que serán analizados para cada uno de los editores o revistas de datos, siendo identificado cada uno de los elementos por un número identificador que será utilizado posteriormente.

Tabla I. Elementos analizados en las revistas de datos

Id	Elemento	Id	Elemento
1	Título	19	Agradecimientos
2	Autores	20	Contribuciones del autor
3	Afiliaciones	21	Conflicto de intereses
4	Palabras Clave	22	Figuras
5	Resumen	23	Leyendas de Figuras
6	Dedicatoria	24	Tablas
7	Tabla de Contenidos	25	Tablas adicionales
8	Antecedentes	26	Convenciones Visuales
9	Especificaciones	27	Referencias
10	Enlace directo a depósito de datos	28	Citaciones
11	Introducción	29	Descripción Conjuntos de Datos
12	Métodos	30	Opinión
13	Diseño Experimental	31	Comentarios del artículo
14	Materiales	32	Resultados
15	Análisis	33	Discusión
16	Registro de Datos	34	Conclusiones
17	Validación Técnica	35	Apéndices
18	Opcional uso de notas	36	Criterio de evaluación

A partir de este conjunto de elementos se tabuló la información encontrada: en cada fila, se representa un editor; en cada columna, uno de los elementos mencionados y fue marcada, en el cruce de cada fila y columna, con una X la existencia de este elemento en las revistas del respectivo editor.

En las figuras 3 y 4 pueden observarse ejemplos de Artículos de Datos publicados por algunas revistas pertenecientes a los editores analizados. El primer ejemplo (ver figura 3) corresponde a la revista *PhotoKeys* del editor Pensoft. En él se identifican claramente el tipo de artículo como *data paper*, la estructura o elementos que componen este artículo.

En la figura 4, se muestra un Artículo de Datos identificado como *data descriptor* de la revista *Science Data* del editor Nature, donde también se encuentran los elementos que conforman este artículo.

Continuando con el análisis de frecuencia de los elementos de la Tabla I en cada uno de los editores analizados, la Tabla II presenta los elementos que presentaron mayor frecuencia para los editores (superior o igual a 7) y la Tabla III, aquellos cuya frecuencia fue baja (menor o igual a 6). De la Tabla III fueron eliminados aquellos editores que no presentaron ningún elemento con bajas frecuencias.

En la Tabla III se presentan los elementos con frecuencias inferiores a 7 que no fueron registrados. Estos elementos corresponden a aquellos que presentaron frecuencias cuyo valor fue 1 (6, 7, 17, 21, 30, 31, 35 y 36), ya que se consideraron secciones o elementos particulares del editor por la especificidad en la mayoría de los casos de la revista y/o por la adaptación a partir de un artículo tradicional. Asimismo, los elementos con frecuencia 2 tampoco fueron considerados relevantes (9, 16, 18 y 25) por las mismas razones. Por lo tanto, el número final de elementos analizados fue 24.

Es importante mencionar que durante la revisión se encontró que la editorial académica independiente Pensoft¹ ofrece a los autores una herramienta informática llamada Integrated Publishing Toolkit (IPT) desarrollada por Global Biodiversity Information Facility (GBIF²). A partir de una plataforma de software, esta herramienta permite a los autores generar de forma automática un Artículo de Datos tras el registro de los metadatos del conjunto de datos que se desea publicar. La mayoría de los metadatos solicitados paso a paso son obligatorios y las figuras y tablas solo podrán ser añadidas como parte del conjunto de datos y no como elementos del Artículo de Datos.

Después de obtener los 24 elementos de análisis, fueron revisadas guías y documentos (Dedieu, 2014; Penev y otros, 2011; Kunze y otros, 2011; Gray, 2015; Chavan, 2011) que forman parte del estado del arte y de la práctica de los Artículos de Datos. En ellos, se encontraron descritos principalmente metadatos que permitían representar de manera completa un conjunto de datos. Es decir, hacían referencia a detalles específicos de los datos como formato, idioma y metodología de muestreo. Esta información no fue detallada por muchos de los editores de las revistas y para este trabajo fue considerada importante y complementaria a los elementos analizados previamente.

Posteriormente, se construyó la propuesta de estructura que debería tener un Artículo de Datos como Publicación a partir de la identificación de cada uno de los elementos que la componen y donde puede notarse que la mayoría de la información de un Artículo de Datos se genera desde los metadatos del conjunto de datos de investigación. Cada

elemento presentado en la Tabla IV tiene su origen de la Tabla II que presentó las mayores frecuencias de los elementos analizados junto con el análisis de otros documentos mencionados anteriormente.

Los elementos «Leyendas de figuras» y «Convenciones visuales» (elementos 23 y 26, respectivamente) con frecuencias superiores a 7 no se encuentran incluidos como secciones de la propuesta, porque se consideran aspectos de formato que facilitan el entendimiento o lectura de un documento. Sin embargo, pueden y deberán ser tenidos en cuenta durante la elaboración de un Artículo de Datos. Por otra parte, el «Enlace directo al depósito de datos» (elemento 10) estaba presente de manera explícita sólo para 3 editores y, en la Tabla IV, está

incluido como parte de la «Descripción del conjunto de datos» debido a que los identificadores de los conjuntos de datos tienen asociada la URL que indica dónde se encuentran los datos en la Web. Además, las otras referencias consultadas que se interrelacionaron para obtener esta propuesta lo consideran un metadato necesario del conjunto de datos.

En la Tabla IV, también se presenta para cada elemento una breve descripción que contempla los metadatos que deben ser considerados por cada elemento y que fueron obtenidos de las guías y documentos mencionados en esta sección. La terminología utilizada forma parte del vocabulario del ámbito científico con el cual se encuentran familiarizados los investigadores.

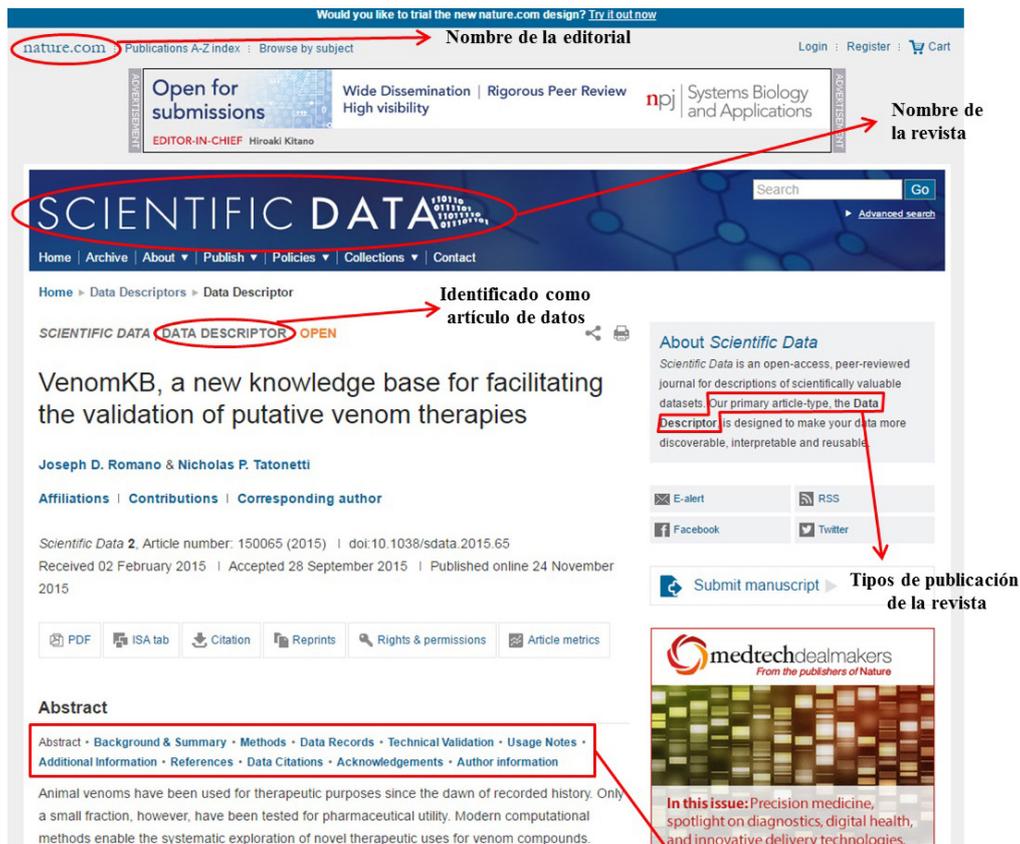
Figura 3. Ejemplo Artículo de Datos revista *PhytoKeys* de la editorial Pensoft

The image shows a screenshot of the PhytoKeys website. Red annotations highlight several key elements:

- Nombre de la editorial:** PENSOFT (circled in red).
- Nombre de la revista:** PhytoKeys (circled in red).
- Identificado como artículo de datos:** Data Paper (circled in red).
- Elementos del artículo de datos:** A table of contents on the right side of the article page, listing sections like Abstract, Keywords, Project details, Taxonomic coverage, Spatial coverage, Methods, etc.
- Tipos de publicación de la revista:** A list of publication types in the left sidebar, with 'Data Papers' checked (circled in red).

Fuente: <http://phytokeys.pensoft.net/articles.php?id=5482>

Figura 4. Ejemplo Artículo de Datos revista *Scientific Data* de la editorial Nature Publishing Group



Fuente: <http://www.nature.com/articles/sdata201565>

Tabla II. Elementos que se presentan con mayor frecuencia en los Artículos de Datos, según el editor

	ELEMENTOS													
	1	2	3	5	11	12	19	22	23	24	26	27	28	29
EDITORES														
BIOMED CENTRAL	X	X	X		X	X		X		X		X		
CHEMISTRY CENTRAL	X	X	X					X		X	X		X	X
COPERNICUS GMBH	X	X	X	X	X		X	X	X	X	X	X		X
ECOLOGICAL SOCIETY OF AMERICA	X	X		X	X	X	X	X		X	X		X	X
ELSEVIER	X	X	X	X	X	X		X	X	X		X		
F1000 RESEARCH LTD	X	X						X	X	X				X
HINDAWI	X	X	X	X	X	X	X	X	X	X		X	X	X
NATURE PUBLISHING GROUP	X	X	X	X	X	X	X	X	X	X		X	X	
PENSOFT	X	X	X	X		X	X						X	X
PLOS	X	X	X	X	X	X	X	X	X	X	X	X	X	X
WILEY	X	X			X			X		X	X	X	X	
SAGE PUBLICATIONS	X	X	X	X	X		X	X	X	X	X	X		
SPRINGER OPEN	X	X	X					X		X	X		X	X
UBIQUITY PRESS	X	X	X	X	X		X	X		X	X	X	X	
TOTAL FRECUENCIA	14	14	11	9	10	7	8	13	7	13	8	9	9	8

Tabla III. Elementos que se presentan con baja frecuencia en los Artículos de Datos, según el editor

		ELEMENTOS									
		4	8	10	13	14	15	20	32	33	34
EDITORES	BIOMED CENTRAL		X						X	X	X
	COPERNICUS GMBH		X					X	X		
	ECOLOGICAL SOCIETY OF AMERICA	X				X			X	X	X
	ELSEVIER			X	X	X	X	X			
	F1000 RESEARCH LTD						X		X		
	HINDAWI		X	X	X						
	NATURE PUBLISHING GROUP		X				X	X			
	PENSOFT	X	X	X	X	X					
	PLOS					X		X	X	X	X
	SAGE PUBLICATIONS	X									
	UBIQUITY PRESS							X			
TOTAL FRECUENCIA		3	5	3	3	4	3	5	5	3	3

Tabla IV. Estructura propuesta para un Artículo de datos

ELEMENTO	DESCRIPCIÓN
Título	Nombre asociado al conjunto de datos.
Autores	Nombre completo de los autores de las contribuciones y proceso de recolección y gestión de los datos de investigación.
Afiliaciones	Datos de la(s) organización(es) a la(s) cual(es) se encuentra adscrito cada autor, ciudad, país, email y dirección de correspondencia.
Resumen	Breve descripción del contenido del artículo y de los datos.
Introducción	Presenta el contexto general y específico de la investigación, la(s) pregunta(s) de investigación que conllevaron a la recolección de los datos, el valor y/o potencial de los datos.
Métodos	Asociado a los métodos y condiciones (de tiempo, espacio, etc.) de recolección y reproducción de los datos, metodología del muestreo, materiales utilizados y guía para la reutilización de los datos.
Descripción del conjunto de datos	Descriptores físicos, formatos, estructura de los datos, idioma, dominios o rango de valores que pueden tomar los datos, explicaciones de datos con valores especiales, identificador del conjunto de datos que puede ser el DOI, depósito de los datos (enlace y tipo).
Figuras, tablas.	Soporte para la presentación de información relacionada con los datos como análisis estadísticos, validación de muestras, entre otros.
Agradecimientos	Reconocimientos a personas, organizaciones, entre otros que el autor considere.
Citaciones	Enlace o citación a artículo(s) tradicional(es) publicado(s) a partir del conjunto de datos por los autores u otros en caso de que estos existan.
Referencias	Lista de las referencias a literatura citada en el texto del artículo.

Finalmente, los elementos que conforman esta propuesta son aquellos que facilitarán la reutilización de los datos y el reconocimiento de los autores. Los resultados tabulados, producto del análisis de frecuencia y los documentos relacionados que existen en la actualidad, permitieron inicialmente tener un conjunto amplio de elementos o metadatos derivados de las aproximaciones y demás propuestas previas que aún no se estandarizan y que, en su mayor parte, obedecen a una adaptación del artículo convencional y particularidades que cada editor añade y que inclusive pueden no ser relevantes para los posteriores investigadores que reutilicen los datos. Los 11 elementos que componen la estructura son de carácter necesario para garantizar una estructura común y homogénea. Además, un aspecto importante durante la revisión fue determinar que, en realidad, los elementos propuestos existen para todos los conjuntos de datos recolectados por los investigadores y forman parte del proceso y las propiedades de Publicación de Datos. También, se asocian a los metadatos o información relevante de los datos. Es importante aclarar que existen más elementos, como los encontrados en las revistas y fuentes analizadas (ver Tabla I), que no forman parte de la estructura propuesta, pero que podrán añadirse a un Artículo de Datos específico de una revista. Para este trabajo se consideran opcionales.

6. CONCLUSIONES

La estructura propuesta en este trabajo para un Artículo de Datos contiene aquellos elementos necesarios que permitirán la compartición y Publicación de datos para que puedan ser reutilizados por otros. Además, contar con una estructura de referencia facilitará el proceso de revisión por pares expertos y las validaciones técnicas necesarias bajo un conjunto de elementos o parámetros definidos. Esto, a su vez, posicionará al Artículo de Datos como una publicación de carácter científico con el mismo rigor y reconocimiento que poseen actualmente los artículos tradicionales o convencionales en los que se presentan resultados de investigación. Esta estructura se compone de metadatos del conjunto de datos de investigación publicados. El Artículo de Datos tiene implícito que los datos serán no solo compartidos, sino que deberán ser publicados obedeciendo todo lo que el proceso de Publicación de datos conlleva.

Por otra parte, un Artículo de Datos permitirá a los autores el reconocimiento no sólo de los resultados obtenidos, sino también del proceso de recolección y publicación que permitirá la reproducibilidad de la ciencia. Otros investigadores, a través

del Artículo de Datos, entenderán y podrán utilizar de forma adecuada un determinado conjunto de datos que, a su vez, optimizará los recursos físicos y financieros de los proyectos y las actividades de investigación.

Finalmente, a partir de la consolidación de los beneficios y desventajas de la compartición de datos realizadas en este trabajo, es importante proponer posibles soluciones o mejoras desde las Ciencias de la Información que serán oportunidades e ideas de investigación futuras.

7. AGRADECIMIENTOS

Los autores agradecen al Grupo de Inteligencia Computacional (GICO) de la Universidad del Cauca (Colombia) y al Grupo de *Pesquisa Novas Tecnologias em Informação* (GPnti) de la Universidad Estadual Paulista Júlio de Mesquita Filho (Brasil).

ACKNOWLEDGMENTS

The authors acknowledge the Computational Intelligent Research Group (GICO) at the University of Cauca (Colombia) and Research Group in New Information Technologies (GPnti) at the Universidade Estadual Paulista Julio de Mesquita Filho (Brazil).

8. NOTAS

1. <http://pensoft.net/about>
2. <http://www.gbif.org/what-is-gbif>
3. Las guías y directrices mencionadas en este documento que hicieron parte del análisis fueron principalmente los siguientes:
 - Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle (ICPSR, 2012).
 - Cartilha Técnica para Publicação de Dados Abertos no Brasil v1.0. Secretaria de Logística e Tecnologia da Informação – SLTI Ministério do Planejamento Orçamento e Gestão – MP. 2012.
 - Critérios, política e procedimentos para a admissão e a permanência de periódicos científicos na Coleção SciELO Brasil, São Paulo – Setembro 2014

9. REFERENCIAS

- Borgman, C. L. (2012). The Conundrum of Sharing Research Data. *Journal of the American Society for Information Science and Technology*, vol. 63(6), 1059-10. <https://doi.org/10.1002/asi.22634>
- Callaghan, S.; Donegan, S.; Pepler, S.; Thorley, M.; Cunningham, N.; Kirsch, P.; ... Wright, D. (2012). Making data a first class scientific output: Data citation and publication by NERC's environmental data centres. *International Journal of Digital Curation*, vol. 7(1), 107-113. <https://doi.org/10.2218/ijdc.v7i1.218>
- Candela, L.; Castell, D.; Manghi, P.; Tani, A. (2015). Data Journals: A Survey. *Journal of the Association for Information Science and Technology*, vol. 66(9), 1747-1762. <https://doi.org/10.1002/asi.23358>
- Chavan, V.; Penev, L. (2011). The data paper: a mechanism to incentivize data publishing in biodiversity science. *BMC Bioinformatics*, vol. 12(15), 1-12. <https://doi.org/10.1186/1471-2105-12-S15-S2>
- Committee for a Study on Promoting Access to Scientific and Technical Data for the Public Interest, Commission on Physical Sciences, Mathematics, and Applications, Division on Engineering and Physical Sciences, National Research Council (CSPASTDPI; CPSMA; DEPS;NCR). (1999). *A question of balance: Private rights and the public interest in scientific and technical databases*. Washington, D.C.: National Academy Press. Disponible en: <https://books.google.com.br/books?id=D2SdAgAAQBAJ> [Consultado: 28 de Octubre de 2015]
- Dedieu, L. (2014). Rédiger et publier un data paper dans une revue scientifique en 5 points. *CIRAD*, Disponible en: <http://url.cirad.fr/ist/data-paper>. [Consultado: 2 de Noviembre de 2015]
- García-García, A.; López-Borrull, A.; Peset, F. (2015) Data journals: eclosión de nuevas revistas especializadas en datos. *El profesional de la información*, vol. 24(6), 845-854. <https://doi.org/10.3145/epi.2015.nov.17>
- Gray, S. (2015). *Case study: Publishing a data paper*. Universidad de Bristol: Research Data Service. Disponible en: <https://data.bris.ac.uk/files/2015/05/Publishing-a-data-paper.pdf>. [Consultado: 2 de Diciembre de 2015]
- Inter-university Consortium for Political and Social Research (ICPSR). (2012). *Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle* (5 ed.). Ann Arbor, Michigan: Inter-university Consortium for Political and Social Research doi:<http://dx.doi.org/10.3886/GuideToSocialScienceDataPreparationAndArchiving>
- Kratz, J; Strasser, C. (2014). Data publication consensus and controversies [version 1; referees:1 approved with reservations]. *F1000Research*, vol. 3(94). <https://doi.org/10.12688/f1000research.3979.1>
- Kunze, J. A.; Cruse, P.; Hu, R.; Abrams, S.; Hastings, K.; Mitchell, C.; Schiff, L.R. (2011). Practices, Trends, and Recommendations in Technical Appendix Usage for Selected Data-Intensive Disciplines. *California Digital Library*. UC Office of the President: California Digital Library. Disponible en: <http://escholarship.org/uc/item/9jw4964t>. [Consultado: 25 de Octubre de 2015]
- Lawrence, B.; Jones, C.; Matthews, B.; Pepler, S.; Callaghan, S. (2011). Citation and peer review of data: Moving towards formal data publication. *International Journal of Digital Curation*. vol. 6(2), 4-37. <https://doi.org/10.2218/ijdc.v6i2.205>
- National Information Standards Organization (NISO). (2013). *RP-15-2013: Recommended Practices for Online Supplemental Journal Article Materials*. Baltimore, MD: National Information Standards Organization. Disponible en: http://www.niso.org/apps/group_public/download.php/10055/RP-15-2013_Supplemental_Materials.pdf [Consultado: 19 de Mayo de 2016].
- National Institutes of Health (NIH). (2003). *NOT-OD-03-032: Final NIH Statement on Sharing Research Data*. MD, USA: National Institutes of Health Bethesda. Disponible en: <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html> [Consultado: 2 de Diciembre de 2015].
- National Science Board; National Science Foundation (NSB;NSF). (2005). *Long-lived digital data collections: Enabling research and education in the 21st century*. Washington, D.C: National Science Foundation (U.S.). Disponible en: <http://www.nsf.gov/pubs/2005/nsb0540/start.jsp>. [Consultado: 5 de Noviembre de 2015]
- Newman, P.; Corke, P. (2009). Editorial: Data papers - peer reviewed publication of high quality data sets. *The international journal of robotics research*, vol. 28(5), 587. <https://doi.org/10.1177/0278364909104283>
- Organisation for Economic Co-operation and Development (OECD). (2007). *OECD Principles and Guidelines for Access to Research Data from Public Funding*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264034020-en-fr>
- Penev, L.; Mietchen, D.; Chavan, V.; Hagedorn, G.; Remsen, D.; Smith, V.; Shotton, D. (2011). *Pensoft Data Publishing Policies and Guidelines for Biodiversity Data*. Pensoft Publishers. Disponible en: http://www.pensoft.net/J_FILES/Pensoft_Data_Publishing_Policies_and_Guidelines.pdf. [Consultado: 14 de Noviembre de 2015]
- Sant'Ana, R. (2013). *Ciclo de vida dos dados e o papel da Ciência da Informação*. XIV Encontro Nacional de Pesquisa em Ciência da Informação, ENANCIB; Florianópolis, Brasil. Disponible en: <http://enancib2013.ufsc.br/index.php/enancib2013/XIVenancib/paper/view/284/319>.
- Tenopir, C.; Allard, S.; Douglass, K.; Aydinoglu, A.U.; Wu, L.; Read, E.; Manoff, M.; Frame, M. (2011) Data Sharing by Scientists: Practices and Perceptions. *PLoS ONE*, vol. 6(6): e21101. <https://doi.org/10.1371/journal.pone.0021101>
- Torres.-Salinas, D. (2009). Compartir datos (data sharing) en ciencia: contexto de una oportunidad. *Anua-*

- rio ThinkEPI*, vol. 1, 258-261. Disponible en: <https://sites.google.com/site/torressalinas/>. [Consultado: 12 de Noviembre de 2015]
- Torres-Salinas, D.; Robinson-García, N.; Cabezas-Clavijo, Á. (2012). Compartir los datos de investigación: introducción al data sharing. *El profesional de la información*, vol. 21(2), 173-184. <https://doi.org/10.3145/eipi.2012.mar.08>
- Van den Eynden, V.; Bishop, L. (2014). *Incentives and motivations for sharing research data, a researcher's perspective*. UK Data Archive, University of Essex: Knowledge Exchange Disponible en: http://repository.jisc.ac.uk/5662/1/KE_report-incentives-for-sharing-researchdata.pdf. [Consultado: 20 de Octubre de 2015]
- Wallis, J.C.; Rolando, E.; Borgman, C. L. (2013). If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. *PLoS ONE*, vol. 8(7): e67332. <https://doi.org/10.1371/journal.pone.0067332>
- Whyte, A.; Callaghan, S.; Tedds, J.; Mayernik, M.S. (2013). Perspectives on the role of trustworthy repository standards in data journal publication. *International Association for Social Science Information Services and Technology Conference*. Cologne, Germany. Disponible en: <http://www.iassistdata.org/conferences/2013/presentation/3657>. [Consultado: 10 de febrero de 2016]