

Abordajes de metadatos para publicadores de datos gubernamentales en el contexto de Iberoamérica*

Resumen

Generar valor para la sociedad a partir de datos gubernamentales se ha tornado ineludible mediante estrategias que los divulgan en conjuntos de datos. Los conjuntos de datos constituyen un ejemplo de reunión de bases de datos que pretenden obtener éxito al establecerse como catálogos desde el punto de vista de los ciudadanos, y amplían la visibilidad de y sobre las acciones de la gestión pública. El objetivo de esta investigación fue describir la adherencia al uso de metadatos en conjuntos de datos de gobiernos de países, con base en el contexto y el marco conceptual presentados en las instrucciones para publicadores de datos encontradas en los sitios de datos abiertos de los países analizados. Se trata de una investigación descriptiva, enfocada en la observación crítica de los documentos que abordan el tratamiento descriptivo de conjuntos de datos gubernamentales. Se utilizan como procedimientos la revisión bibliográfica y documental y la definición de estudios de caso en Colombia, Brasil, España y Portugal. Fue posible determinar el ahínco de los gobiernos al abordar con preferencia la categoría metadatos en el uso, lo que da así mayor importancia al aspecto práctico en relación con el teórico, sin desconocer la relevancia de las instrucciones teóricas. En el análisis de contenido se observaron como destacadas las instrucciones provistas por el Gobierno de España, en las que el empeño por aproximar a los usuarios está orientando la publicación de esos documentos; ese país aportó la mayor cantidad de documentos disponibles, mayor número de categorías abordadas y mayor intensidad en el trato otorgado a cada componente temático observado. Se sugiere que los gobiernos ejecuten planes para mejorar sus orientaciones a publicadores de datos a partir de herramientas, tanto conceptuales como prácticas, y de documentos prolijos con componentes teóricos y pragmáticos para el tratamiento descriptivo de conjuntos de datos.

Palabras clave: metadatos, representación de datos, datos de gobierno, publicadores de datos, conjuntos de datos.

Cómo citar este artículo: Camperos-Reyes, Jacquelin Teresa; Sant'Ana, Ricardo César Gonçalves (2021). Abordajes de metadatos para publicadores de datos gubernamentales en el contexto de Iberoamérica. *Revista Interamericana de Bibliotecología*, 44(3), e339110. <https://doi.org/10.17533/udea.rib.v44n3e339110>

Recibido: 2019-07-12 / **Aceptado:** 2021-09-01

* El presente texto publica los resultados finales de la disertación de maestría, estudio que permitió obtener el grado de magíster en Ciencia de la Información por la Universidade Estadual Paulista -UNESP. Asimismo, es un avance del proyecto “Ciclo de Vida dos Dados: impactos e alternativas para redução de assimetrias informacionais” del Grupo de Investigación Tecnologías de Acceso a Datos - GPTAD de la Universidade Estadual Paulista - UNESP, Marília - Brasil.

Jacquelin Teresa Camperos-Reyes

Doctoranda en Ciencia de la Información en la Universidade Estadual Paulista (UNESP), Marília - Brasil. Magíster en Ciencia de la Información de la UNESP, Marília - Brasil. Ingeniera de Sistemas de la Universidad Francisco de Paula Santander (UFPS), Cúcuta - Colombia.

jacquelin.camperos-reyes@unesp.br
<https://orcid.org/0000-0002-0078-5376>

Ricardo César Gonçalves Sant'Ana

Doctor en Ciencia de la Información por la Universidade Estadual Paulista (UNESP), Marília - Brasil. Magíster en Ciencia de la Información de la UNESP, Marília - Brasil. Graduado em Matemática y Pedagogía. Profesor del Programa de Pós-Graduação en Ciencia de la Información de la UNESP, Marília - Brasil. Profesor Asociado de la UNESP, Facultad de Ciencias e Ingenierías (FCE), Tupã - Brasil.

ricardo.santana@unesp.br
<https://orcid.org/0000-0003-1387-4519>



Metadata Approaches for Government Data Publishers in the Ibero-American context

Abstract

Generating value for society from government data has become unavoidable through strategies that make them available in datasets. The datasets are an example of gathering databases that aim to be successful by standing as catalogs from the point of view of citizens, expanding the visibility of and on actions of public management. The objective is to describe the adherence to the use of metadata in datasets of governments of countries, based on the context and conceptual framework presented in the instructions for publishers of data found in open data sites of the countries analyzed. It is a descriptive investigation, focused on the critical observation of documents that address the descriptive treatment of government datasets. The bibliographic and documentary review and the definition of case studies in Colombia, Brazil, Spain, and Portugal are used as procedures. It was possible to determine the zeal of the data retainers by preferentially addressing the category metadata in use, thus giving greater importance to the practical aspect in relation to the theoretician, without disregarding the relevance of the theoretical instructions. In the content analysis, the instructions provided by the Government of Spain were observed as outstanding, where the effort to bring users closer to each other is guiding the publication of these documents; that country provided the largest number of documents available, the largest number of categories addressed, and the greatest intensity in the treatment given to each thematic component observed. It is suggested that governments implement plans aimed at improving their orientation to data publishers by providing both conceptual and practical tools, offering lengthy documents with theoretical and pragmatic components for the descriptive treatment of datasets.

Keywords: Metadata, data representation, government data, data publisher, dataset.

1. Introducción

Uno de los desafíos de la ciencia de la información es transformar la actual abundancia de datos, de modo que permita, además de la toma de decisiones, añadir valor a las actividades realizadas en un determinado entor-

no (Santos, 2010). Así, se hace necesario considerar la estructuración de los datos como recursos informacionales, en este caso datos de gobierno, como un proceso basilar en la búsqueda de ampliación y eficiencia en el acceso y uso de esos datos.

McGarry (1999) asegura que los recursos informacionales deben ser estructurados de alguna forma para evitar su amorfismo y su inutilidad; incluso, Cornella (2010) asigna el término *infoxicación* al actual fenómeno de la excesiva producción de información, que no en pocas ocasiones permanece amorfa. Como afirma Aleixo (2020), transitamos por unos estados de anomia, entre ellos la anomia informacional, entendida como la “ausencia de reconocimiento y de la práctica de normas que se centren hacia la elaboración de estructuras que reglamenten un proceso de acciones que busquen una patronización, para incluir y estimular la participación de la sociedad” (Aleixo, 2020, p. 119).

En consecuencia, Sant’Ana (2013), frente a la abundancia en la oferta de flujos informacionales digitales, señala que la decisión de desarrollar medios innovadores de representación de esos contenidos contribuye a la recuperación y acceso a esos recursos informacionales, y coadyuva en la condición primaria de convivencia y supervivencia para el fortalecimiento de las capacidades humano-informacionales.

En este sentido, los conjuntos de datos son datos tabulados con cierta estructura, definidos por la comunidad Schema.org como una colección de recursos informacionales estructurados que “[...] describen uno o varios temas de interés” (Schema, 2017, para. 1).

El caso de los conjuntos de datos gubernamentales constituye un ejemplo de reunión de bases de datos que pretenden obtener éxito al estructurarse como catálogos centrales desde el punto de vista de los ciudadanos; estos amplían la visibilidad de y sobre las acciones de la gestión pública y son “una forma inteligente de medir y monitorear la salud de las bases públicas disponibles para la sociedad” (Governo do Estado São Paulo, 2015, p. 23).

Algunos estudios se han publicado en torno a la estructuración de esas bases de datos gubernamentales. Por ejemplo, los autores Moreira, Zafalon, Santos y

Sant'Ana (2017), a partir de la pregunta-problema cuáles son los elementos de metadatos que los servicios de disponibilización de datos gubernamentales están utilizando para representar sus recursos informacionales, esto es, en el proceso de recuperación de datos a partir del sitio de datos del gobierno brasileño, examinaron de forma cruzada los elementos de metadatos usados en la publicación frente al esquema de metadatos creado para la representación de datos del gobierno de ese país, el patrón Estándar de Metadatos para el Gobierno Electrónico (e-PMG).

Otro estudio, de los autores Camperos-Reyes, Sant'Ana y Santarém-Segundo (2018), analizó conjuntos de datos de agricultura y desarrollo rural, para determinar si esos datos están siendo, o no, publicados como datos abiertos a la luz de los principios de Sebastopol; uno de los principios observados instruye sobre datos estructurados pasibles de ser procesados por máquinas.

Por otro lado, Silva, Soares, Cesar y Rabelo (2020) examinaron datos publicados por ministerios del gobierno de Brasil bajo tres dimensiones: inteligibilidad, operacionalidad e interactividad y observaron, en cada dimensión, la calidad de los datos publicados. La dimensión que trata sobre el aprovechamiento de los datos desde el punto de vista de la estructura con que son *disponibilizados* fue la nombrada como *operacionalidad*, al concluir que hay un alto grado de no-operacionalidad debido a problemas como la publicación de archivos no estructurados, archivos inaccesibles, enlaces inoperantes, estandarización deficiente y problemas de integridad, que señalan el alto costo de la apropiación de esos datos publicados, costo asumido por los usuarios de los sitios.

Así pues, el objetivo de este estudio fue describir la adherencia al uso de metadatos en conjuntos de datos de gobiernos de algunos países, tomando como base el contexto y el marco conceptual presentados en las instrucciones para publicadores de datos, encontradas en los sitios de datos abiertos de los países analizados.

Considerando que los metadatos son elementos que permiten representar las características de los recursos informacionales, atendiendo sus atributos de contenido, contexto y estructura (Gilliland, 2008), y que de su correcto uso se pueden derivar beneficios en su orga-

nización, interactividad, identificación y preservación (Santos, Simionato & Arakaki, 2014), el foco de la investigación está en documentos de apoyo disponibles en sitios de datos abiertos de Colombia, Brasil, España y Portugal, los cuales podrían contener instrucciones para la publicación de conjuntos de datos, examinando lo relacionado con el uso de metadatos dentro del proceso descriptivo de esos recursos de información.

2. Metodología

Se adoptó una investigación descriptiva, utilizando como procedimientos metodológicos para la recolección de los datos la revisión bibliográfica y el análisis documental, así como la aplicación de estudios de caso por países para delinear el escenario concerniente a la representación de recursos informacionales, en este caso, conjunto de datos gubernamentales.

El volumen de datos e información obtenido durante la recolección fue abordado mediante la técnica de análisis de contenido (Bardin, 2010), sobre la base de que el proceso de interpretación es fundamentalmente interactivo, en el cual “[...] el investigador elabora poco a poco una explicación lógica del fenómeno o de la situación estudiados, examinando las unidades de sentido, las interrelaciones entre esas unidades y entre las categorías en que se encuentran reunidas” (Gil, 2002, p. 90).

En los estudios de caso se deben redoblar los cuidados tanto en la planificación y recolección como en el análisis de los datos (Gil, 2002), razón por la cual se usó el análisis de contenido como instrumento que otorgue el rigor científico deseado para la consecución del objetivo general de la investigación.

El análisis de contenido puede aplicarse, entre otras fuentes, a unidades textuales, tanto en comunicaciones formales como informales. Farias et al. (2020) utilizaron esa técnica para examinar algunas respuestas obtenidas mediante formularios en línea aplicados a profesionales de la ciencia de la información. En otro caso, Moreira, Valentim y Sant'Ana (2018) analizaron contenido sobre los planes de enseñanza de universidades brasileñas e hicieron inferencias acerca de la interdisciplinariedad de la ciencia de la información y sus contribuciones para el estudio de la divulgación de datos de gobierno en inter-

net. Por su parte, França y Carvalho (2019) encontraron que en investigaciones sobre redes sociales y bibliotecas, publicadas en periódicos científicos brasileños, el análisis de contenido es la técnica más usada como instrumento para el análisis de los datos recolectados.

Para este estudio, se desarrolló un análisis que permitiera inferir el panorama sobre las indicaciones que dirigen el uso de metadatos durante el proceso de publicación de conjunto de datos en los sitios de datos abiertos de los gobiernos de Brasil, Colombia, España y Portugal.

Para el desarrollo del análisis de contenido, se realizaron, según lo prescrito por Bardin (2010), estas tres fases cronológicas: 1) preanálisis, 2) exploración del material y 3) tratamiento de los resultados.

2.1 Fase de preanálisis

De acuerdo con la propuesta de Bardin (2010) en lo que se refiere al privilegio que tienen los procedimientos exploratorios en la fase de preanálisis, el descubrimiento de los recursos que conformaron el corpus de esta investigación comienza con la definición de la muestra; se optó por considerar, en un primer momento, la realidad del continente americano, con referencia a condiciones de países del Cono Sur. Estos países presentan similitudes históricas fundamentadas en sus procesos de colonización, además, en la disponibilidad de sus recursos naturales y en la forma en que esos países se insertan en la economía mundial (Lucci, Blanco & Mendonça, 2013).

Más allá de la realidad de América del Sur, algunos países europeos presentan una relación tanto histórica como cultural y política con la realidad local y, por lo tanto, se incluyeron como foco en este estudio. Para la muestra de países se seleccionaron Colombia, Brasil, España y Portugal.

Los sitios de datos gubernamentales para los estudios de caso son, respectivamente, Colombia: <https://datos.gov.co/>; Brasil: <http://dados.gov.br/>; España: <http://datos.gob.es/>; y Portugal: <http://www.dados.gov.pt/pt/inicio/inicio.aspx>

En cada sitio se exploraron los sistemas de organización presentes, y se consideró que la información

relacionada de hecho se dirige a una audiencia específica. Con ello, se navegó en los esquemas de organización y se consideraron a) tópicos, b) orientación a tareas y c) esquemas para audiencias específicas (Rosenfeld, Morville & Arango, 2015).

Con base en la revisión bibliográfica y el análisis documental, el criterio de búsqueda de los documentos en cada sitio fue recuperar todos los recursos que tuvieran como propósito instruir al capital humano que participa de la publicación de conjunto de datos, con orientaciones sobre los mecanismos de representación, metadatos, de los mismos conjuntos de datos.

Con la adopción de la propuesta de Bardin (2010) para la constitución del corpus de la investigación, se observaron las siguientes reglas en la distinción de los documentos: a) exhaustividad, b) representatividad, c) homogeneidad y d) pertinencia.

Una vez reunido el conglomerado de documentos, este se depuró para identificar los documentos que efectivamente contienen instrucciones relacionadas con la descripción de conjuntos de datos.

2.2 Fase de exploración del material

Para la administración de las técnicas en el corpus (Bardin, 2010, p. 128), se realizó la exploración del material depurado y se eligió como unidad de registro el análisis temático, asumiendo que “hacer un análisis temático consiste en descubrir los núcleos de sentido que componen la comunicación y cuya presencia, o frecuencia de aparición, pueden significar algo para el objetivo analítico escogido” (Bardin, 2010, p. 131). El tema ofrece una unidad de recorte de sentido y no de forma, “[...] que no es suministrada de una vez por todas, ya que el recorte depende del nivel de análisis y no de manifestaciones formales reguladas” (Bardin, 2010, p. 131).

Para ello, el tema eje para la exploración fue metadatos, que actúan como objeto referente en esta fase del análisis. En los documentos recuperados se identificaron fragmentos en los que los autores abordan el tratamiento descriptivo de los conjuntos de datos, teniendo en cuenta las variaciones de los idiomas español y portugués (metadatos y *metadados*).

Una vez definido el tema, que actuaría como unidad de registro, se consideraron las categorías para clasificar los abordajes identificados en los documentos, esto a partir del estudio desarrollado por Santos et al. (2013).

Se adoptaron dos de las categorías determinadas por el estudio de Santos et al. (2013), usadas aquí para identificar el tipo de abordaje dado a la temática metadatos en los documentos: teoría y uso. El abordaje de metadatos se clasificó en la categoría teoría cuando estos se trataron a partir del “examen y de la observación generando conceptos, fundamentos, técnicas, estudios, explicaciones y críticas” (Santos et al., 2013, p. 3); y se asignaron a la categoría uso cuando se explicitaron “experimentos, relatos de experiencia, estudios de caso, recomendaciones, y percepciones de uso” (Santos et al., 2013, p. 3).

Las nomenclaturas usadas para sistematizar este análisis fueron las siguientes:

- Metadatos teoría (MT): conceptos, MT (C); fundamentaciones, MT (F); técnicas, MT (T); estudios, MT (E); explicaciones, MT (EX); y críticas MT (CR).
- Metadatos en el uso (MNU): experimentos, MNU (EX); relatos de experiencia, MNU (RE); estudios de caso, MNU (EC); recomendaciones, MNU (R); y percepciones de uso MNU (PU).

Fue necesario establecer reglas de enumeración o modo de conteo (Bardin, 2010), y para este análisis fueron adoptadas dos reglas: la presencia o ausencia de las categorías arriba citadas y la intensidad con que se trata la temática, según variaciones semánticas o formales. Así, para la evaluación de la intensidad en el abordaje

de las unidades de registro, se consideró la intensidad de los verbos estimando los adverbios que los pueden acompañar, así como el énfasis y la profundización de los sintagmas en los que se inserta la unidad de registro. La unidad de contexto de la técnica del análisis fue cada documento recuperado con instrucciones.

Para ayudar en la caracterización de los documentos que fueron analizados se creó un esquema que integró las categorías utilizadas en la codificación de las unidades de registro. El esquema incluye atributos que describen los documentos recuperados, las categorías, las reglas de conteo y los esquemas de metadatos indicados para los publicadores de datos.

Luego de la fase de exploración del material, los datos brutos se examinaron de modo que pudieran ser significativos en la consecución de los objetivos de esta investigación.

2.3 Fase de tratamiento de los resultados

¿Cuáles son las significaciones de los mensajes contenidos en los documentos? ¿Cuáles son los contenidos de la comunicación en los documentos analizados? Asimilando la teoría de Bardin (2010), ¿cuáles podrían ser las subcategorías presentes según la propuesta de la codificación de los documentos y la profundización percibida?

La variable de inferencia giró en torno al objetivo general de esta investigación, con ella se buscó observar la adherencia al uso de los metadatos en los documentos para usuarios de gobiernos involucrados en la publicación de conjuntos de datos. La síntesis de la técnica aplicada se puede observar en la Tabla 1.

Tabla 1. Síntesis de la técnica aplicada en la investigación

| Preanálisis | Exploración del material | Tratamiento de los resultados |
|--|---|---|
| Exploración de los sitios por país para encontrar recursos: documentos con instrucciones para publicadores de datos. | Unidad de registro: análisis temático. Tema: metadatos/metadados | Polo que sería observado: mensaje en los documentos, respecto al significado (Bardin, 2010). |
| Reglas de distinción de los recursos: <ul style="list-style-type: none"> ● exhaustividad; ● representatividad; ● homogeneidad; ● pertinencia (Bardin, 2010). | Reglas de enumeración: <ul style="list-style-type: none"> ● presencia o ausencia de las categorías; ● intensidad, con base en variaciones semánticas o formales de las categorías (Bardin, 2010). | Variable de inferencia: adherencia al uso de metadatos en los documentos que contienen instrucciones de los órganos de gobiernos para los involucrados en la publicación de conjuntos de datos. |

Fuente: elaboración propia.

2.4 Datos sobre los países de la muestra: Colombia, Brasil, España y Portugal

En el contexto mundial de publicación de datos de gobierno, Colombia ha tenido un buen desempeño en lo relacionado con la implementación de estrategias encaminadas a impulsar la publicación de datos a partir del gobierno nacional (Organización para la Cooperación y el Desarrollo Económico [OCDE], 2017; Global Editors Network, 2017). En el sitio de datos de ese gobierno, www.datos.gov.co, actualmente están publicados 5965 conjuntos de datos (Gobierno de Colombia, 2021), provenientes de diversos órganos del Estado colombiano.

Por su lado, Brasil no solo es el país sede de esta investigación. Es uno de los pioneros de América Latina y del mundo en la promoción y ejecución de políticas de Estado, con repercusión y convocatoria para otros países, que atienden anhelos de la sociedad en la percepción de transparencia del gobierno, mediante la utilización y aprovechamiento de la coyuntura ofrecida por la internet para la disponibilidad y uso de datos abiertos de gobierno.¹

La OCDE (2017) desarrolló *The OCED OURdata Index*, un marco de trabajo con indicadores para mensurar la apertura de datos de gobiernos. El índice permite sintetizar puntos fuertes y débiles, identificando áreas de acción que puedan influenciar las políticas nacionales en países miembros, o candidatos a serlo, de ese organismo transnacional.

Tanto el gobierno de Portugal como el de España hicieron parte de la primera convención de la OCDE en 1960, realizada en la busca de la prosperidad y el fortalecimiento económico de países, así como el incremento de la calidad de vida en ellos.

En los casos de Colombia y Brasil, hasta el momento, ninguno es miembro pleno de esa organización. El pedido de adhesión de Colombia fue admitido en 2013, siendo considerado por la OCDE como un país con proceso de adhesión “en estado de discusión”; mientras que Brasil, ante la propia manifestación de interés, espera por el inicio de la discusión para adhesión. Sin embargo, para la OCDE, Brasil es considerado como un aliado

1 Ver <https://www.opengovpartnership.org/about/about-ogp>

clave, particularidad que comparte con India, Indonesia, China y África del Sur (OCDE, 2018).

En la última medición del *OURdata Index*, que recolectó datos en 2019, se observaron esfuerzos desarrollados por los gobiernos para la promoción, uso, reutilización y puesta a disposición de herramientas para la publicación de datos. En esa edición, Colombia aparece en tercer lugar en el ranking global, que tiene en sus primeros puestos a Corea del Sur y Francia (OCDE, 2019).

El índice desarrollado por la OCDE considera tres aspectos: disponibilidad de los datos, accesibilidad a los datos y soporte gubernamental para la reutilización de los datos. Este índice tiene una magnitud que va de 0 a 1, calculado según la metodología diseñada por la propia OCDE (2017). La Figura 1 presenta el *OURdata Index* para los cuatro países de la muestra.

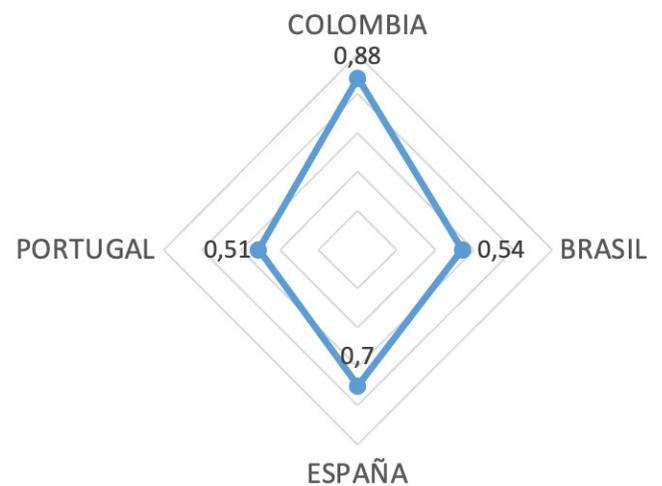


Figura 1. *OURDATA Index* en países de la muestra.

Nota: el índice de Brasil corresponde a la medición publicada en el año 2017. En la versión publicada en 2019 no hay datos en ese índice para ese país.

Fuente: adaptado de OCDE (2017, 2019).

Como se indicó, el desempeño de Colombia fue destacado, a pesar de ser uno de los países en proceso de adhesión. Se observa mayor puntuación en relación con los otros países, clasificados en orden decreciente: Colombia, España, Brasil y Portugal.

Una vez adoptada la metodología, la colecta de documentos con instrucciones en Colombia, Brasil, España

y Portugal fue llevada a cabo entre noviembre y diciembre de 2017.

3. Resultados

Finalizado el análisis de contenido en las instrucciones recuperadas por cada país, a continuación, son presentados los resultados en cada una de sus fases.

3.1 Fase de preanálisis

Observando el sistema de organización del sitio datos.gov.co de Colombia, se encontró, en la ruta del menú Herramientas – Aprender sobre datos – Guías y Manuales, documentación relacionada con publicación y descripción de conjunto de datos. De los 11 documentos encontrados, 4 se relacionan con el objetivo de la investigación.

En el caso de Brasil en el sitio dados.gov.br, el menú inferior contiene la opción *Manuais* (Manuales), desde el que fueron seleccionados 2 de los 14 documentos disponibles por cumplir con los criterios indicados en la metodología.

La navegación por el menú principal del sitio datos.gob.es de España, en la ruta Interactúa – Documentación

– Guías, permitió seleccionar 14 documentos que cumplieron los objetivos del estudio.

Finalmente, en el sitio dados.gov.pt de Portugal, se observó que a pesar de no haber una opción específica para documentación sobre procesos de publicación de conjuntos de datos, un documento divulgado en la sección de noticias aborda esta temática.

La síntesis de los documentos que conforman el cuerpo del estudio está presentada en la [Tabla 2](#).

Los cuatro países, en la fecha de la recolección, tenían publicados 41 documentos que posiblemente estaban relacionados con instrucciones para la publicación de conjuntos de datos. Sin embargo, tras el análisis, de esos 41 documentos solo 14 contenían efectivamente orientaciones para la descripción de esos conjuntos de datos.

3.2 Exploración del material

Se realizó la gestión del corpus de la investigación aplicando de forma sistemática la observación de los fragmentos en que se abordó el tratamiento descriptivo de los conjuntos de datos, considerando los términos metadatos y *metadados*, por la variación en las lenguas castellana y portuguesa.

Tabla 2. Documentos seleccionados en la fase de preanálisis

| Colombia | Brasil |
|--|---|
| <ul style="list-style-type: none"> • Guía de datos abiertos en Colombia (Gobierno de Colombia, 2016a); • Guía para la estandarización de datos abiertos (Gobierno de Colombia, 2016b); • Manual de datos abiertos en Colombia para publicador (Gobierno de Colombia, 2016c); • Manual de datos abiertos en Colombia para administrador (Gobierno de Colombia, 2016d). | <ul style="list-style-type: none"> • Cartilha Técnica para Publicação de Dados Abertos no Brasil (Gobierno de Brasil, 2012); • Manual de Catalogação no Portal Brasileiro de Dados Abertos (Gobierno de Brasil, 2016b). |
| España | Portugal |
| <ul style="list-style-type: none"> • Manual de usuario: perfil organismo público (Gobierno de España, 2017a); • Plantilla RDF/XML para la federación de conjuntos de datos (Gobierno de España, 2017b); • Orientaciones y garantías en los procedimientos de anonimización de datos personales (Gobierno de España, 2016b); • Orientaciones sobre la protección de datos en la reutilización de la información del sector público (Gobierno de España, 2016a); • Guía de aplicación de la Norma Técnica de Interoperabilidad (Gobierno de España, 2016c); • Como publicar datos abiertos de manera rápida y sencilla (con CKAN) (Gobierno de España, 2015); • Guía de aplicación del Real Decreto 1495/2011 (Gobierno de España, 2012). | <ul style="list-style-type: none"> • Guia Dados abertos (Gobierno de Portugal, 2016). |

Fuente: elaborado propia a partir de los datos recolectados.

[Abordajes de metadatos para publicadores de datos gubernamentales en el contexto de Iberoamérica]

Con respecto a la codificación, para el tratamiento del corpus, se observaron las categorías definidas a partir del estudio de Santos et al. (2013), con el fin de clasificar y organizar los documentos analizados, y examinar “el eslabón entre los datos de los textos y la teoría” (Bardin, 2010, p. 129).

Las reglas de enumeración para esta etapa, como se indicó en la metodología, corresponden a la presencia o ausencia de las categorías metadatos teoría y metadatos en el uso. La intensidad de estos enfoques se observó según la vehemencia semántica de los sintagmas en los elementos del análisis, es decir, de las categorías adoptadas.

3.3. Tratamiento de los resultados

Bardin (2010) indica que los resultados brutos deben tratarse de manera que sean significativos y válidos. Para ello, el resultado de la exploración puede ser observado de manera sintética mediante figuras (gráficas en las secciones subsecuentes) y retornando a las unidades del estudio de casos.

3.3.1 Colombia

La Figura 2 compendia la categorización de los documentos analizados en el contexto colombiano, en ella se asumen las siguientes siglas para referenciar los cuatro documentos recuperados: a) DOC 1: *Guía de datos abiertos*

en Colombia (Gobierno de Colombia, 2016a); b) DOC 2: *Guía para la estandarización de datos abiertos* (Gobierno de Colombia, 2016b); c) DOC 3: *Manual de datos abiertos en Colombia para publicador* (Gobierno de Colombia, 2016c); y, d) DOC 4: *Manual datos abiertos en Colombia para administrador* (Gobierno de Colombia, 2016d).

En el eje horizontal están referenciados tanto la identificación de los documentos analizados en Colombia como las categorías observadas en el análisis de contenido. En el eje vertical se representan las intensidades en el abordaje de las categorías según alta (A), media (M) o baja (B).

En el corpus explorado predomina el abordaje de la categoría MNU. Esto se sustenta en que en todos los documentos recuperados se discuten indicaciones prácticas de uso que los publicadores de datos deben atender como parte del proceso descriptivo de los conjuntos de datos. Sin embargo, es notable que solo uno de los documentos se manifiesta con alto nivel de intensidad.

Así mismo, al observar el sentido de la unidad de registro, dos documentos abordan la categoría MT, y se obtiene en ambos casos un tratamiento con nivel de intensidad bajo. Se trata de los documentos *Guía de datos abiertos en Colombia* (Gobierno de Colombia, 2016a) y *Guía para la estandarización de datos abiertos* (Gobierno de Colombia, 2016b).

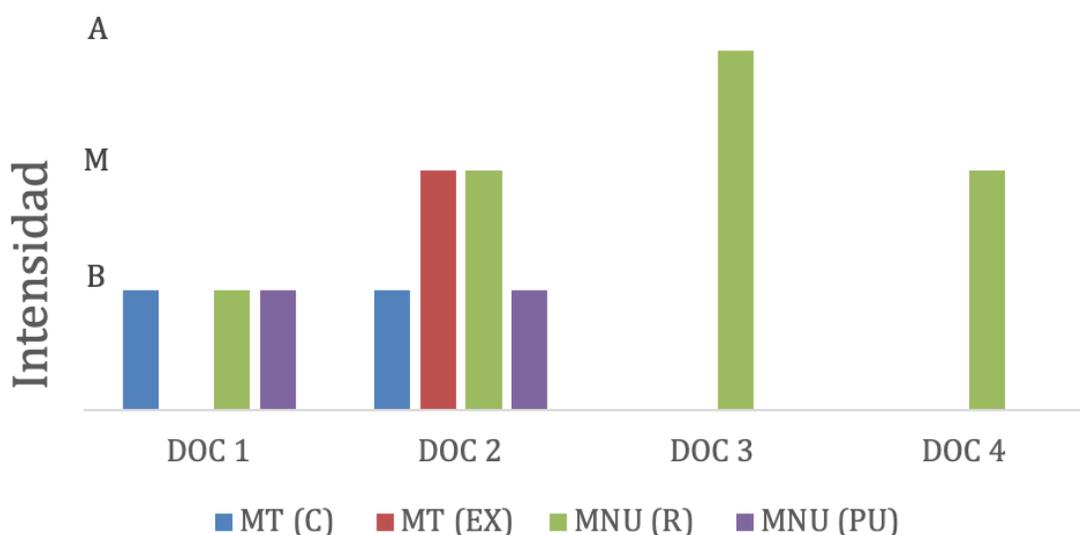


Figura 2. Resultados Colombia.

Fuente: elaborado propia a partir de los datos recolectados.

En el análisis sobre la variable de inferencia, se percibió en las intensidades observadas que el corpus de investigación abarca de manera predominante las recomendaciones, MNU (R), en lo relacionado con el proceso de descripción de los conjuntos de datos, es decir, hay una mayor intensidad en los contenidos que abordan la categoría MNU, con respecto a recomendaciones de uso de esos datos que están siendo publicados.

Estas recomendaciones en el uso de metadatos apuntan a la observación de directrices técnicas en el contexto de publicación de datos abiertos, específicamente, las directrices y recomendaciones del consorcio World Wide Web (W3C).² Colombia indica explícitamente acoger el fundamento producido por ese consorcio. En el documento *Guía para la estandarización de datos abiertos*, en la sección “Metadatos”, se indica el uso del vocabulario *Data Catalog Vocabulary* (DCAT) como referencia para la descripción de los conjuntos de datos (Gobierno de Colombia, 2016b).

El documento arriba mencionado destaca la importancia de la descripción apropiada de los conjuntos de datos para garantizar el éxito en el acceso y la recuperación de datos. Además, se refiere a los metadatos como elementos enriquecedores de los datos contenidos en los conjuntos de datos y destaca un esquema en el que es posible observar la conexión de los metadatos con los objetivos que estos deben alcanzar durante el proceso de acceso a los conjuntos de datos y a las distribuciones asociadas a ellos (Gobierno de Colombia, 2016b).

Los beneficios señalados por el Gobierno colombiano con respecto a esas recomendaciones de uso de metadatos están relacionados con interoperabilidad; economía de esfuerzos en proyectos de software; uso y reutilización de datos, y estrategias de gestión y seguimiento mediante sistemas de indicadores estandarizados (Gobierno de Colombia, 2016b).

El documento *Manual de datos abiertos en Colombia para publicador* enfatiza en la amplia y suficiente descripción de los metadatos que los usuarios publicadores deben registrar durante el proceso de publicación. El documento trae una descripción extensiva del proceso. Esto lo hizo destacarse en la categoría MNU, en el tópico re-

2 Ver <https://www.w3.org/TR/vocab-dcat/>

comendaciones, MNU (R), en el que se percibe como con intensidad alta (Gobierno de Colombia, 2016c).

En la categoría MNU, solo un documento aborda la percepción de uso de los metadatos (PU). Es el caso del documento *Guía de datos abiertos en Colombia*, el cual evidencia la posibilidad de, al tener datos correctamente descritos, referenciarlos o enlazarlos con otros sitios de Colombia que publiquen también datos de gobierno (Gobierno de Colombia, 2016a).

En ese sentido se destacan, aunque sin profundidad, casos de sitios de datos abiertos en países que utilizaron el vocabulario DCAT, así como otros esquemas para el tratamiento descriptivo de los conjuntos de datos, presentando modelos usados por los Estados Unidos y España, y que llegan a describir el uso del perfil de aplicación del vocabulario DCAT, DCAT-AP.

Para la Comisión Europea, el desarrollo de un perfil de aplicación como el DCAT-AP reviste importancia al visualizar la interoperabilidad entre las diversas fuentes de datos de los sitios en el contexto europeo. DCAT-AP es el perfil de aplicación que utiliza el sitio de datos de la Comisión Europea (Portal Europeo de Datos, 2017).

La documentación de Colombia ejemplifica, aunque de manera sucinta, el DCAT-AP como alternativa para un desarrollo futuro que podría traer mayor efectividad en el uso del sitio de datos abiertos de ese país (Gobierno de Colombia, 2016b).

Asimismo, fue posible observar el documento *Manual de datos abiertos en Colombia para administrador*, que presenta la unidad de registro de temática con nivel medio de intensidad, mediante recomendaciones sobre el uso de los metadatos, y señala la posibilidad de personalización de algunos de ellos según la necesidad del publicador (Gobierno de Colombia, 2016d).

Conviene subrayar la falta de precisión lingüística en las instrucciones colombianas al usar la palabra *metadato* para referirse a los atributos de las interfaces gráficas. Este evento podría llevar a errores en las interpretaciones por parte de los usuarios de los documentos. El caso fue observado en los documentos *Manual de datos abiertos en Colombia para publicador* y *Manual de datos abiertos en Colombia para administrador* (Gobierno de Colombia, 2016c, 2016d).

Otra imprecisión se presenta en los sintagmas en los que el documento *Guía para la estandarización de datos abiertos* menciona los términos *metadatos descriptivos* y *metadatos estructurales*. Se observó que el documento relaciona metadatos descriptivos con el conjunto de metadatos que describe cada conjunto de datos, y como metadatos estructurales con los conjuntos de metadatos que van a describir las distribuciones de cada conjunto de datos.³

La categoría analizada MT tuvo un menor conteo en la regla presencia; solo dos documentos la abordaron, ambos con intensidad baja en los tópicos conceptos, MT (C), y explicaciones, MT (EX). En el caso del primer documento categorizado en MT, *Guía de datos abiertos en Colombia*, el tratamiento descriptivo de los conjuntos de datos se aborda bajo el concepto de *documentación de los datos* (Gobierno de Colombia, 2016a, p. 21) para describir los aspectos básicos de los conjuntos de datos frente a la necesidad de “clasificar y mejorar los resultados de la búsqueda y recuperación de los datos” (Gobierno de Colombia, 2016a, p. 21).

En este tópico, se percibe consenso con la definición de metadatos de Alves y Santos (2013) cuando afirman que ellos son “datos que describen otros datos en un sistema de información, con el fin de identificar de forma única una entidad (recurso informacional) para su posterior recuperación” (p. 43). Sin embargo, se percibe que el enfoque teórico expuesto en el documento colombiano necesita una mayor profundidad que proporcione claridad a los usuarios que van a hacer uso de ese documento.

Aunque el documento *Guía para la estandarización de datos abiertos* presenta a los usuarios una descripción detallada de todos los términos de las entidades utilizadas en el modelo de datos DCAT, se observa ausencia de descripción de las relaciones y cardinalidades presen-

tes entre las entidades del modelo. Debe considerarse como importante esta descripción con el fin de aproximar adecuadamente los usuarios a los modelos de datos (Gobierno de Colombia, 2016b).

Se percibe, además, que no se ofrece a los usuarios de los documentos *Manual de datos abiertos en Colombia para publicador* y *Manual de datos abiertos en Colombia para administrador* ninguna aproximación teórica o algún modelo de datos en el que sea posible visualizar alguna abstracción del dominio de los conjuntos de datos (Gobierno de Colombia, 2016c, 2016d).

En lo que se refiere a saber si los documentos con instrucciones para los publicadores de datos indican específicamente los metadatos que serán utilizados en la descripción de los conjuntos de datos, y considerando la imprecisión del lenguaje percibido en el uso del término *metadatos descriptivos*, el gobierno de Colombia recomienda la utilización del vocabulario⁴ DCAT, modelo de datos creado por W3C. En la representación gráfica del vocabulario que propone W3C, es posible percibir que además de metadatos propios de DCAT se utilizan metadatos de varios sistemas de representación como Dublin Core, Simple Knowledge Organization System (SKOS)⁵ y Friend of a Friend (FOAF)⁶ (Colombia, 2016b).

Respecto a la elección de Colombia en lo que se refiere a las directrices para los procesos descriptivos de los conjuntos de datos asentados en el W3C, vale la pena aclarar que dicho consorcio recomienda el uso de metadatos descriptivos, “patrones de estructuras de datos”, tanto en los catálogos como en los conjuntos de datos y en sus distribuciones.

3 Este enfoque expuesto en el documento colombiano se podría asociar a los “metadatos descriptivos” y “metadatos estructurales” descritos por Gilliland (2008). La diferencia radica en que en el documento de Colombia la descripción del recurso es en el primer caso el conjunto de datos en sí mismo y, en el segundo, es la descripción de la distribución del conjunto de datos. Gilliland (2008) propone el uso de “patrones de valores de los datos”. Se observa en el contenido del documento *Guía para la estandarización de datos abiertos* (Gobierno de Colombia, 2016b) que este tipo de patrón es llamado simplemente como valores de los metadatos o *identificadores*, en el caso de datos de carácter geográfico.

4 W3C utiliza el término *vocabulario* en lugar de patrón de metadatos para referirse a esquemas recomendados en la catalogación de recursos informacionales como los conjuntos de datos. Ver <https://www.w3.org/TR/vocab-dcat/>

5 Vocabulario, recomendado por el W3C, que provee un modelo de datos para la disponibilización y distribución de sistemas de organización del conocimiento en la web. SKOS permite representar taxonomías, tesauros, listas de asunto, esquemas de clasificación y otras formas de organización del conocimiento.

6 Vocabulario para uso en internet. Provee metadatos descriptivos respecto a personas y organizaciones (<http://lov.okfn.org/>).

Para los valores de los metadatos, “patrones de valores de los datos”, los documentos presentan una propuesta con el uso de categorías para describir asuntos de los conjuntos de datos, proponiendo en un primer momento 22 ítems para los sectores primarios de la economía en Colombia (Gobierno de Colombia, 2016b).

El documento arriba citado sugiere, además, a) la estandarización de valores de los datos correspondientes a las localizaciones geográficas, indicando el uso del *geonames*; b) la estandarización de los nombres de las entidades y órganos oficiales, proponiendo la creación de una taxonomía de órganos públicos (Gobierno de Colombia, 2016b, p. 63); y c) los datos sobre los procesos de calidad que se desarrollaron durante el proceso de publicación de los conjuntos de datos. En este punto, no se direcciona el uso de ningún metadato específico que aborde la calidad en los conjuntos de datos.

3.3.2 Brasil

Los documentos publicados por el sitio de datos brasileño que componen el corpus son a) DOC 1: *Cartilla Técnica para la publicación de datos abiertos en Brasil* (Gobierno de Brasil, 2012), documento fundamental que contiene el contexto inicial sobre la temática de datos abiertos y directrices básicas dirigidas a órganos de gobierno, todo dentro del marco de la ley de acceso a

la información en Brasil; b) DOC 2: *Manual de catalogación en el portal brasilero de datos abiertos* (Gobierno de Brasil, 2016b), desarrollado para apoyar órganos del poder ejecutivo federal y demás instituciones públicas, específicamente en procesos de catalogación de datos abiertos.

Realizado el análisis de contenido, se llegó a la síntesis presentada en la Figura 3, la cual mantiene la organización de los ejes.

Llama la atención que, de los 14 documentos disponibles en el sitio de datos, solo dos contenían efectivamente instrucciones para publicadores de datos. Se percibió en ese punto que varios de los documentos dispuestos en el sitio contemplan de manera concomitante aspectos relacionados con la gestión de conjuntos de datos, o sea, el contenido de los recursos en ocasiones se presenta repetitivo o referenciando contenido ya abordado en los documentos representados en la Figura 3.

Cada documento analizado aborda de manera exclusiva una de las categorías propuestas en este estudio. En la regla de enumeración presencia, se tiene entonces una aparición en lo que se refiere a la temática MT (C), y una aparición en la temática MNU (PU). Pudo inferirse que ambas categorías fueron abordadas con un nivel de intensidad bajo.

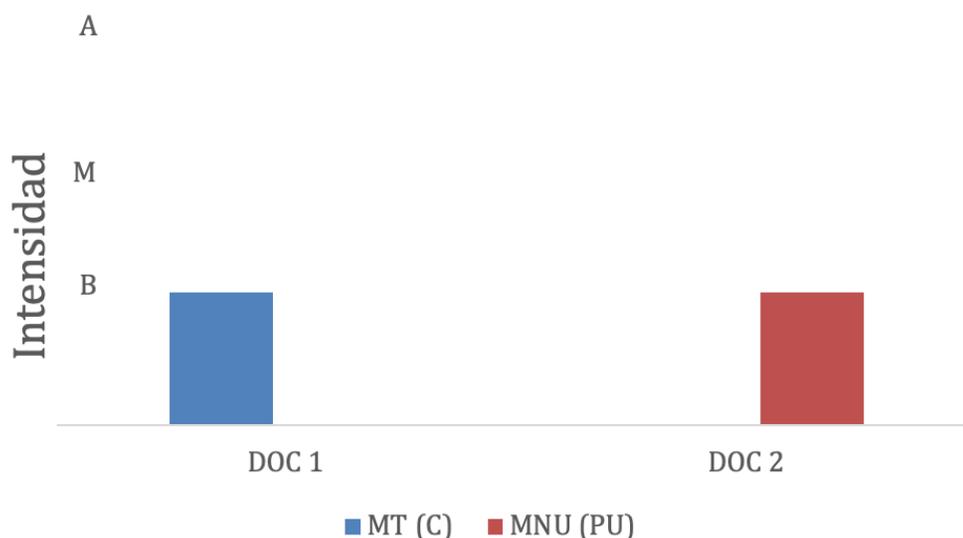


Figura 3. Resultados Brasil.

Fuente: elaborado propia a partir de los datos recolectados.

[Abordajes de metadatos para publicadores de datos gubernamentales en el contexto de Iberoamérica]

Desde el punto de vista de las categorías observadas, es un panorama equilibrado, que sin embargo podría abordarse con un mayor nivel de profundización en lo que se refiere a las fundamentaciones, ejemplificaciones, mejores orientaciones prácticas, de modo que los usuarios publicadores puedan observar y aproximarse a las mejores técnicas de descripción.

También fue posible observar que no está siendo explícitamente indicado el uso de un esquema de metadatos estructural y que, además, presenta divergencias en lo relacionado con el uso de elementos de metadatos indicados en el proceso descriptivo de los conjuntos de datos. Este hecho lo ilustra la indicación para el uso de la plataforma Comprehensive Knowledge Archive Network (CKAN),⁷ en la que, sin embargo, no se pudo observar aproximación explícita hacia alguna estructura de metadatos para la descripción de conjuntos de datos en dicha plataforma.

Los dos documentos de la muestra contienen instrucciones para la inclusión de elementos metadatos; no obstante, con diferencias entre uno y otro documento, como se observa en la variación del orden de inclusión de los metadatos, lo que podría generar confusión para los usuarios publicadores.

Para ilustrar, DOC 1 (Gobierno de Brasil, 2012) hace una clara distinción entre dos tipos de metadatos para la descripción de conjuntos de datos: “obligatorios” y “deseables”. Los metadatos obligatorios incluyen los siguientes elementos: Título, Descripción, Catálogo Origen, Órgano responsable, Categoría del Vocabulario Controlado del Gobierno Electrónico (VCGE)⁸ y Recursos (Gobierno de Brasil, 2012, p. 7).

A su vez, DOC 2 (Gobierno de Brasil, 2016b) indica que, al momento de la creación de conjuntos de datos, los elementos que deben ser considerados en la catalogación, sin distinguir obligatoriedad, son Título,

7 Plataforma gratuita que permite crear sitios para la publicación de datos abiertos, plataforma *open source*. Desarrollada y promovida por la Open Knowledge Foundation (OKFN) (OKFN, 2017): ver <https://ckan.org/> y <https://okfn.org/opendata/how-to-open-data/>

8 VCGE es una lista jerárquica de términos con asuntos del gobierno, usada para indexar informaciones en el gobierno federal y que busca fortalecer la interfaz con la sociedad y proporcionar una herramienta de gestión (Gobierno de Brasil, 2016a, 2016b).

Descripción, Etiquetas, Licencia, Organización y Visibilidad.

Se observa en este escenario que, en la Cartilla, DOC 1, la intención del gobierno brasileño es la correcta descripción de los recursos informacionales, aproximándose a una estrategia descriptiva propia, y que ya existe dentro del contexto del gobierno electrónico. Por otro lado, el Manual, DOC 2, presenta elementos descriptivos pertenecientes a la plataforma CKAN, que es la utilizada por el sitio de datos del Gobierno de Brasil. Se explica en este punto el porqué de las sutiles diferencias en el orden e indicación de los elementos de metadatos entre estos dos documentos.

Otro hecho que sustenta la inferencia arriba indicada es la orientación de incluir de forma obligatoria el metadato “categoría”, del VCGE, informada en la Cartilla DOC 1 (Gobierno de Brasil, 2012, p. 7), situación que no se especifica con tal grado de obligatoriedad en el DOC 2.

Otro mecanismo federal creado para fortalecer el entorno del gobierno electrónico en Brasil es el e-PMG. El e-PMG fue fundamentado en el estándar Dublin Core, manteniendo para sí los 15 elementos básicos del estándar internacional, añadiendo cinco elementos más, considerados necesarios en el contexto del gobierno electrónico brasileño (Gobierno de Brasil, 2014).

Se observó que en los documentos recuperados y analizados no hay mención sobre el estándar e-PMG, hecho que puede suponer que, a pesar de que el Gobierno de Brasil haya trabajado con un mecanismo de representación para sus recursos informacionales, no está siendo vinculado a la plataforma de datos abiertos de gobierno.

Una buena práctica observada en las instrucciones brasileñas es que durante el proceso de catalogación se menciona un diccionario de datos que permite describir las columnas en las distribuciones de los conjuntos de datos (Gobierno de Brasil, 2016b). Aunque el Manual no utiliza el término *metadatos* en ese punto, es claramente una indicación de la importancia manifiesta de un proceso descriptivo en el contexto estudiado.

El elemento que el sitio de datos brasileño nombra como *diccionario de datos* se corresponde con el sitio de datos colombiano, allí denominado *metadatos estructurales*. Es-

tos dos términos se aplican para describir la estructura en las distribuciones de los conjuntos de datos.

3.3.3 España

Este país aportó un número significativo de documentos, siete en total, en los cuales se identificaron seis de las categorías temáticas.

La nomenclatura adoptada en la [Figura 4](#) para identificar los documentos en el eje horizontal es a) DOC 1: *Manual de usuario: perfil organismo público* (Gobierno de España, 2017a); b) DOC 2: *Plantilla RDF/XML para la federación de conjuntos de datos* (Gobierno de España, 2017b); c) DOC 3: *Orientaciones y garantías en los procedimientos de anonimización de datos personales* (Gobierno de España, 2016b); d) DOC 4: *Orientaciones sobre la protección de datos en la reutilización de la información del sector público* (Gobierno de España, 2016a); e) DOC 5: *Guía de aplicación de la Norma Técnica de Interoperabilidad* (Gobierno de España, 2016c); f) DOC 6: *Cómo publicar datos abiertos de manera rápida y sencilla* (con CKAN) (Gobierno de España, 2015); y,

g) DOC 7: *Guía de aplicación del Real Decreto 1495/2011* (Gobierno de España, 2012).

Las dos grandes categorías, MT y MNU, aparecen con tres subcategorías cada una.

Se destaca el hecho de que dos de las tres temáticas abordadas en la categoría MT se registraron con intensidad alta: técnicas, MT (T), y explicaciones, MT (EX). Asimismo, MNU, con las subcategorías recomendaciones, MNU (R) y percepciones de uso, MNU (PU), tienen ocurrencia con alta intensidad: de seis categorías abordadas en las instrucciones españolas, cuatro lo fueron con un alto nivel de detalle.

El abordaje que las instrucciones hacen de las temáticas fue percibido con un nivel de profundidad altamente provechoso para los publicadores de datos. En DOC 5 (Gobierno de España, 2016c), hay una descripción amplia y completa del modelo de datos recomendado para ser utilizado en el tratamiento descriptivo de los conjuntos de datos. El documento contempla tres categorías temáticas, todas con un nivel de intensidad alto: MT (T), MT (EX) y MNU (R).

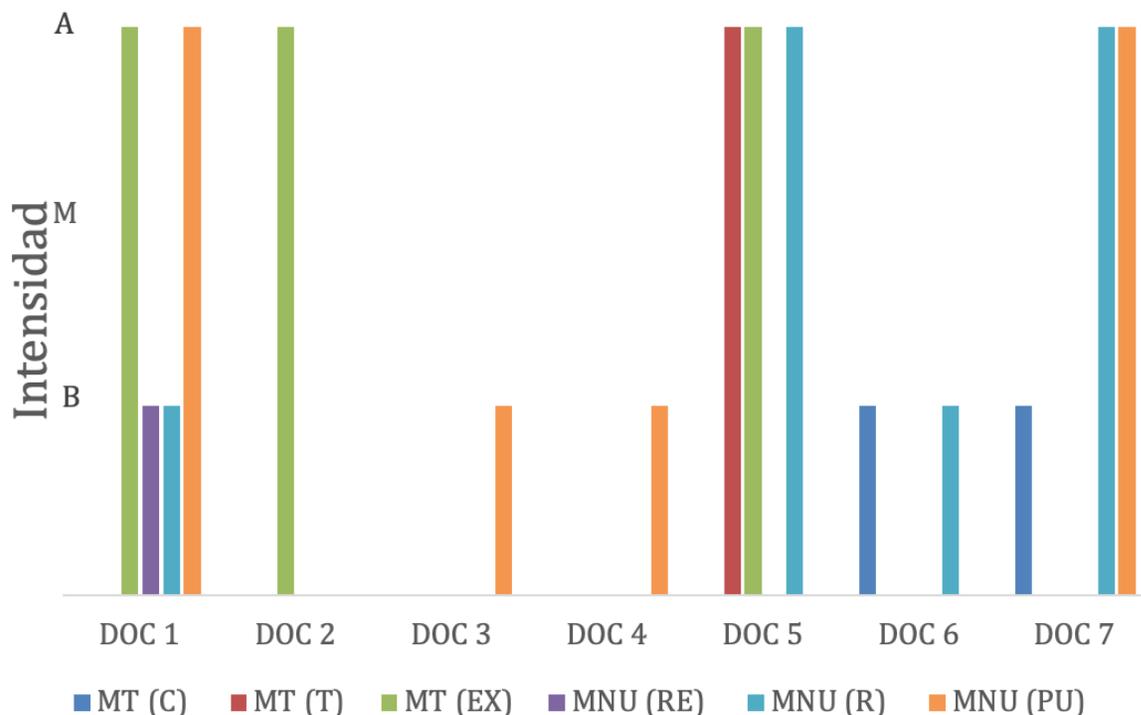


Figura 4. Resultados España.

Fuente: elaborado propia a partir de los datos recolectados.

[Abordajes de metadatos para publicadores de datos gubernamentales en el contexto de Iberoamérica]

Se indica en el documento citado el uso del modelo DCAT y se presenta una descripción de los tipos de relaciones entre entidades, cardinalidad y características de los atributos del modelo recomendado por el Gobierno español. Esta descripción detallada otorga claridad a los usuarios publicadores de datos.

El documento también especifica la inclusión de relaciones de las entidades del modelo DCAT⁹ con esquemas de conceptos SKOS para describir órganos del gobierno, recursos geográficos y taxonomías temáticas. Por otra parte, indica que estos tipos de mecanismos de representación, SKOS y taxonomías, se utilizan como metadatos de valor, es decir, como metadatos descriptivos (Gobierno de España, 2016c).

Otros elementos contenidos en la guía arriba mencionada son la descripción de *namespaces*¹⁰ que utiliza el modelo DCAT y el detalle del diccionario de datos de este modelo de descripción (Gobierno de España, 2016c).

El diccionario de datos describe entidades del modelo DCAT incluyendo elementos de metadatos con sus descripciones, especificando elementos de los modelos FOAF, SKOS y Dublin Core que son usados por el modelo. Además, señala la obligatoriedad de uso de algunos elementos y la posibilidad de múltiples valores, para finalmente indicar el tipo de valores que son permitidos (Gobierno de España, 2016c).

Un recurso que presenta cuatro categorías temáticas, tanto de MT como de MNU, es DOC 1 (Gobierno de España, 2017a); en él se destaca que el contenido se enfoca en mantener equilibrio entre el abordaje teórico y el uso de esas teorías tratadas. En lo que respecta a MNU, el documento explica la posibilidad de que cada órgano publicador incluya relatos de aplicaciones y de reutilización de datos de conjuntos de datos.

Este documento enfatiza, además, en la correcta catalogación de los conjuntos de datos para la posterior generación de estadísticas que permiten retroalimen-

9 Además de DOC 5, otros tres documentos indican el uso del vocabulario DCAT como modelo de datos: (Gobierno de España, 2012, 2017a, 2017b).

10 Alternativa tecnológica que permite distinguir elementos, instancias, entre diferentes vocabularios (ver <http://lov.okfn.org/>).

tar la gestión de la plataforma de datos de España. En el sitio de datos se indican alternativas para observar el uso de los datos publicados por órganos del Estado (Gobierno de España, 2017a).

La documentación de España es la única que tiene dos documentos con instrucciones que abordan la categoría temática MT (EX) con nivel de intensidad alto. Es el caso de DOC 2, que explicita ampliamente la descripción de conjuntos de datos mediante elementos del vocabulario DCAT y orienta al usuario publicador mediante comentarios y observaciones en cada elemento de descripción (Gobierno de España, 2017b).

La documentación española proporciona también documentos con instrucciones para abordar aspectos de privacidad. Mediante el tratamiento de la temática MNU (PU), otorga atención a elementos de metadatos como aspecto determinante que debe ser considerado en el proceso de anonimización y de protección de datos personales. Estos documentos son DOC 3 (Gobierno de España, 2016b) y DOC 4 (Gobierno de España, 2016a).

DOC 4 discute sobre la importancia de los metadatos como elementos que, además de ser entregados en abierto y procesables por máquina, cuenten también “con los niveles más altos de precisión y desagregación” (Gobierno de España, 2016a, p. II). Santos y Sant'Ana (2013) señalan la utilidad de la granularidad con fines de funcionalidad y utilización de los datos y reafirman dicha característica como enriquecedora desde la perspectiva de la ciencia de la información por el impacto directo en el acceso y uso de esos datos publicados.

En cuanto a indicación de metadatos descriptivos, a pesar de no existir especificación sobre algún patrón, se indica el uso de taxonomías para tal fin. Como ejemplo, se señala el uso del elemento `dcat:themeTaxonomy`, que pertenece al modelo DCAT, y que en ese punto dirige a la página <http://datos.gob.es/kos/sector-publico/sector>, la cual contiene valores que hacen referencia a categorías o temas de los conjuntos de datos (Gobierno de España, 2017a).

Además, es perceptible que las organizaciones que están sujetas a la publicación de datos se encuentran normalizadas en la plataforma.

DOC I indica el uso de tres recursos para ser empleados como elementos de contenido en algunos de los elementos descriptivos: órganos publicadores,¹¹ ámbitos geográficos¹² y temas.¹³

Los documentos españoles se perciben enfáticos en lo que se refiere a indicaciones y directrices, de tal manera que sean aplicables en la publicación de cada órgano oficial en sus propios sitios, ya que son directrices que implementan el marco legal del país y hacen énfasis en la reutilización de la información.

En España el proceso se centra en publicadores de todos los órganos de gobierno que ofrecen recursos tanto en el sitio de datos del país como en los sitios de datos de cada instancia de gobierno.

Se percibe unicidad de esfuerzos provenientes de la administración española y se observa un trabajo en conjunto de varios órganos oficiales que tratan directamente asuntos del gobierno electrónico; así, se otorga interdisciplinariedad en el abordaje de los aspectos observados en este estudio.

En general, se observa una mayor intencionalidad, por parte de organismos del Gobierno español involucrados en la publicación de datos, para proporcionar subsidios con indicaciones para distintos tipos de usuarios y procesos asociados a la descripción de conjuntos de datos gubernamentales.

3.3.4 Portugal

La Agência para a Modernização Administrativa (AMA) es un organismo vinculado al Gobierno de Portugal, encargado de la gestión del sitio de datos de gobierno. Luego de la navegación, fue posible percibir que no hay una opción que contenga documentación sobre procesos de publicación de conjuntos de datos. Existen dos recursos documentales en el sitio, los cuales fueron accesibles mediante la divulgación de una noticia de noviembre de 2016, y apenas son accesibles en los enlaces publicados en esa nota informativa. De esos recursos, solamente uno se encuadra en la descrip-

11 Ver <http://datos.gob.es/apidata/catalog/publisher>

12 Ver <http://datos.gob.es/apidata/catalog/spatial>

13 Ver <http://datos.gob.es/apidata/catalog/theme>

ción de conjuntos de datos: *Guia dados abertos* (Gobierno de Portugal, 2016).

En ese documento, apenas una categoría temática es abordada: MT (C), con un nivel de baja intensidad. En el documento, el Gobierno portugués hace una sucinta referencia a metadatos descriptivos como elementos determinantes para la reutilización de la información y para la interoperabilidad de los recursos.

El documento detalla, además, la evolución de modelos de metadatos que vienen siendo utilizados en procesos descriptivos de recursos informacionales, lo que implica procesos de normalización y estandarización en órganos que implementan estrategias de publicación de datos.

La Figura 5 presenta los elementos que el documento considera en esa evolución.



Figura 5. Estados en la evolución de modelo de metadatos, Portugal.

Fuente: Gobierno de Portugal (2016).

En el documento está descrita cada fase con características, ventajas e inconvenientes. Se observa la inclusión del modelo de datos DCAT como opción para la representación de conjuntos de datos de gobierno. A pesar de esa indicación precisa, no se observa ninguna descripción del modelo, entidades, relaciones o cardinalidades que ofrezcan algún detalle al usuario publicador.

Portugal, al igual que Colombia, menciona el perfil de aplicación DCAT-AP, lo contextualiza y lo indica como una opción para implementar que podría contribuir con la interoperabilidad en países del continente europeo.

3.3.5 Inferencias en el grupo de países

La unidad temática objeto de este estudio, metadatos, fue abordada mediante las categorías indicadas previamente: MT y MNU. Se encontraron ocurrencias únicamente en las subcategorías MT (C), (MT) (E), MT (T), (MNU) (R), MNU (PU) y MNU (RE).

La **Tabla 3** permite visualizar estas categorías junto a los niveles de intensidad con que fueron desarrolladas, así como las ocurrencias que surgieron en el análisis.

Tabla 3. Síntesis de intensidades en las categorías

| Categoría - Subcategoría | Niveles de intensidad | | |
|-----------------------------|-----------------------|-------|------|
| | Bajo | Medio | Alto |
| MT. Conceptos | 6 | 0 | 0 |
| MT. Explicaciones | 0 | 1 | 3 |
| MT. Técnicas | 0 | 0 | 1 |
| MNU. Recomendaciones | 3 | 2 | 3 |
| MNU. Percepciones de uso | 5 | 0 | 2 |
| MNU. Relatos de experiencia | 0 | 0 | 1 |
| Totales | 14 | 3 | 10 |

Fuente: elaborado propia a partir de los datos recolectados.

En la **Tabla 3** se percibe que, entre las dos categorías propuestas para la clasificación de los abordajes, seis subcategorías fueron registradas según la primera regla de conteo propuesta, *presencia* o *ausencia*, con tres subcategorías de MT y tres de MNU, lo que manifiesta un

equilibrio entre el tratamiento de los metadatos con enfoque teórico y con enfoque de uso.

Con relación a la segunda regla de conteo, *intensidad*, la mayoría se presentó con un abordaje de intensidad baja. Hay 14 registros de subcategorías con tratamiento leve, que deberían gradualmente ser enriquecidos por bien de los procesos de descripción y posterior publicación de datos; además, es posible notar que 10 de las ocurrencias de las subcategorías encontradas se abordaron considerando profusión de detalles y explicaciones necesarias para capacitar a los publicadores de datos.

Los documentos analizados presentan referencias a las categorías con niveles de intensidad a menudo bajos y medios; sin embargo, hay que destacar que, de las seis categorías, en cinco de ellas al menos un documento fue abordado con intensidad alta.

Es interesante acentuar que, en la relación entre la ocurrencia de las categorías de los cuatro países, fue España el que más categorías abordó en sus documentos, las seis encontradas por este estudio fueron identificadas en los recursos recuperados (ver **Figura 6**).

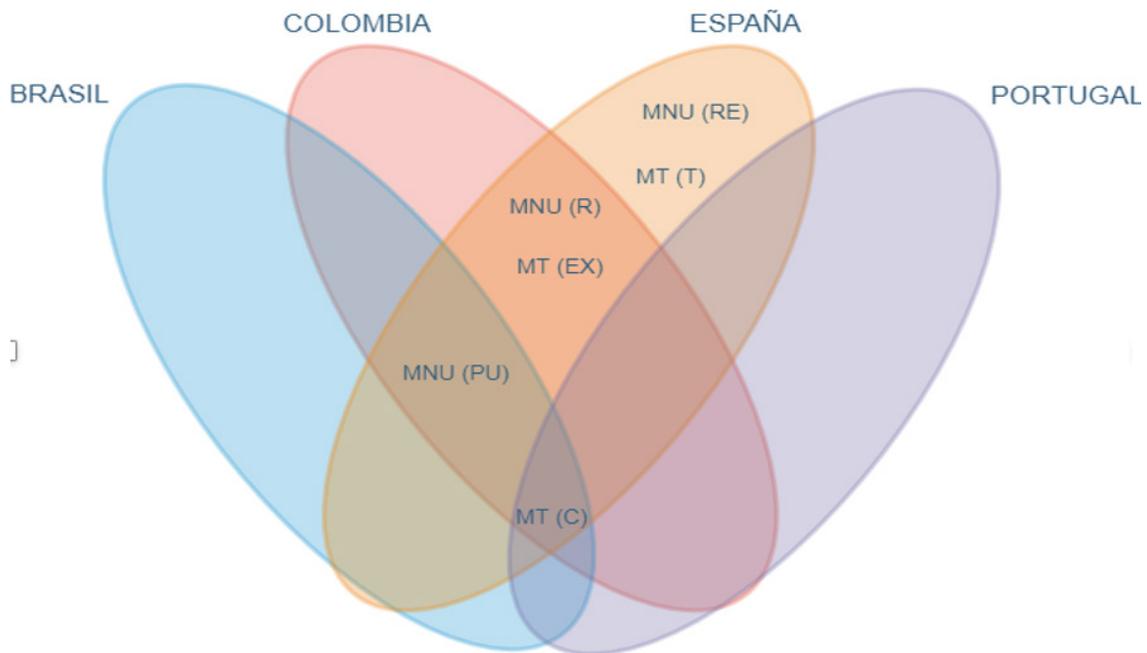


Figura 6. Ocurrencia de las categorías por país.

Fuente: elaborado propia a partir de los datos recolectados.

En lo referente a los documentos españoles, se considera que el empeño por la aproximación a los usuarios está orientando la elaboración de instrucciones al aportar la mayor cantidad de documentos disponibles, mayor número de categorías abordadas y mayor intensidad en el trato otorgado a cada componente temático observado.

Se percibió además la intención del Gobierno español de desarrollar competencias informacionales para usuarios publicadores, no solo al corroborar la alta intensidad en la mayoría de los abordajes realizados, sino también al apreciar el engranaje que el gobierno pretende alcanzar entre diversos órganos oficiales en busca de interoperabilidad de sus recursos informacionales.

Por otra parte, la estrategia descrita en los documentos del sitio de datos de Colombia podría considerar el perfeccionamiento del enfoque conceptual en el tratamiento descriptivo de los conjuntos de datos, con el fin de presentar con mayor claridad dicha información a los usuarios. Sin embargo, se observó una tendencia positiva en los contenidos analizados y se encontraron, en un punto medio en el desarrollo de instrucciones para publicadores de datos, instrucciones que se perciben aceptables al presentar cuatro categorías con intensidades medias y altas.

En cuanto a la falta de precisión lingüística en algunos fragmentos de los documentos analizados en ese país, se recomienda el perfeccionamiento en el uso de términos para que sean pertinentes, para evitar confusión y errores en las interpretaciones por parte de los usuarios. Una posible solución sería la fundamentación de los documentos con referencia teórica, de modo que los usuarios puedan consultar y elucidar lo expuesto en ellos.

Del lado de Brasil pudo observarse que la plataforma de datos de gobierno suministra diversas opciones que pretenden desarrollar competencias de los usuarios publicadores de datos al ofrecer varias opciones documentales, incluso entregando recursos didácticos para acompañar el proceso de publicación.

Sin embargo, se sugiere que estos documentos pasen por procesos de mejoramiento en lo referente al marco teórico y de uso de metadatos, con el fin de entregar herramientas conceptuales y prácticas para mejorar, por un lado, la percepción teórica y, por otro, la descripción

en sí, mediante, por ejemplo, la exposición de casos de éxito y diversos usos del proceso descriptivo de conjuntos de datos.

Por último, se considera que la documentación proporcionada por Portugal, encontrada como escasa, podría ampliarse tanto en cantidad como en la intensidad de los abordajes en las áreas temáticas objeto de este estudio.

Cabe resaltar la importancia que los países hacen en la creación e implementación de perfiles de aplicación, como el caso del DCAT-AP, creado y recomendado por la Comisión Europea.

En la perspectiva de los esquemas de metadatos recomendados en las instrucciones, se observó que Colombia, España y Portugal hacen mención sobre el uso de modelo DCAT para la descripción de los conjuntos de datos. En cuanto a Brasil, no se indica explícitamente la utilización de un modelo de datos. Dicho país opta en la documentación por abordar la descripción de los conjuntos de datos siguiendo únicamente lo que para tal fin contiene la plataforma CKAN.

En cuanto a las plataformas que utilizan los sitios analizados, fue posible constatar que Colombia utiliza Socrata,¹⁴ mientras que Brasil y España optaron por la solución CKAN. En el caso de Portugal, el sitio de datos se ejecuta en una plataforma propia desarrollada y mantenida por la AMA.

4. Consideraciones

Los resultados permiten observar un panorama acerca de la adherencia al uso de metadatos en procesos de publicación de datos de gobierno, considerando documentos con instrucciones para los publicadores. Fue posible determinar el ahínco de los tenedores de datos en los cuatro países analizados al abordar preferentemente la categoría MNU, lo que da así mayor importancia al aspecto práctico del abordaje en relación con el teórico. Sin desconsiderar la relevancia de contar con explicaciones teóricas, el análisis de los documentos identificó que hay más detalles procedimentales rela-

¹⁴ Plataforma para la gestión de datos y servicios en línea dirigida a órganos oficiales como gobiernos de ciudades, estados y países. Socrata es un producto comercial, que requiere el pago de costos relacionados con su implementación y manutención.

cionados con la descripción de conjuntos de datos mediante el uso de metadatos. Hay una mayor cantidad absoluta en las subcategorías abordadas, 16 ocurrencias frente a 11 en la categoría MT.

Se sugiere que los gobiernos ejecuten planes direccionados a mejorar sus orientaciones a publicadores de datos, entreguen herramientas tanto conceptuales como prácticas y ofrezcan documentos prolijos con componentes teóricos y pragmáticos para el tratamiento descriptivo de conjuntos de datos.

En el sentido de la cobertura teórica, que alcanzada con justo equilibrio podría permitir una mejor comprensión de las ventajas de la correcta representación de los conjuntos de datos como recursos informacionales, se entiende la importancia de una amplia y correcta descripción, tanto de las entidades que componen los modelos de datos como de las relaciones y cardinalidades presentes, acercando así a los usuarios a los modelos de datos.

Se observa que la búsqueda del fortalecimiento de las capacidades humano-informacionales (Sant'Ana, 2013) podría pasar por estudios que alcancen uno de los aspectos de ese tipo de capacidades, el registro de los recursos informacionales, a fin de contribuir, como expone Le Coadic (2004), en lo referente al objeto social imperante y necesario de las investigaciones en el área de la ciencia de la información. En este caso particular, como indica este autor, “la demanda acuciante de la tecnología de la información señala caminos de investigación dentro de las prácticas sociales, que estén revestidos de utilidad y eficacia” (Le Coadic, 2004, p. 21).

Son admitidas inquietudes referentes a los procesos de publicación de datos de gobierno y sus relaciones con otros temas de interés, como los socioeconómicos, posibles vínculos con indicadores de desarrollo en países y regiones, así como la reciprocidad de esas estrategias en los grupos sociales y la efectividad transformada en hechos que demuestren el aumento de la transparencia en los gobiernos, todo ello en la perspectiva de investigaciones originadas a partir de la ciencia de la información.

5. Referencias

1. Aleixo, Diana (2020). *O estado de anomia dos dados no acesso aos dados governamentais abertos no Brasil* (tesis de doctorado). Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília. <http://hdl.handle.net/11449/191686>
2. Alves, Rachel; Santos, Plácida (2013). *Metadados no domínio bibliográfico*. Rio de Janeiro: Intertexto.
3. Bardin, Laurence (2010). *Análise de conteúdo*. Lisboa: Edições 70.
4. Camperos-Reyes, Jacquelin Teresa; Sant'Ana, Ricardo César Gonçalves; Santarém, José Eduardo (2018). Estudio comparativo de datasets gubernamentales de Brasil y de Colombia, con datos de Agricultura y Desarrollo Rural. *Anales de Documentación*, 21(2). <https://doi.org/10.6018/analesdoc.21.2.30238>
5. Cornella, Alfons (2010). *Infoxicación: buscando un orden en la información*. Barcelona: Infonomia.
6. Gobierno de Brasil (2012). *Ministério do Planejamento, Orçamento e Gestão. Cartilha técnica para publicação de dados abertos no Brasil v1.0*. <http://dados.gov.br/pagina/cartilha-publicacao-dados-abertos>
7. Gobierno de Brasil (2014). *Padrão de Metadados do Governo Eletrônico - ePMG*. https://www.gov.br/governodigital/pt-br/governanca-de-dados/PMGVersao1_1.pdf
8. Gobierno de Brasil (2016a). *Vocabulário Controlado do Governo Eletrônico - VCGE*. https://www.gov.br/governodigital/pt-br/governanca-de-dados/VCGE_2_1_0.pdf
9. Gobierno de Brasil (2016b). *Ministério do Planejamento, Desenvolvimento e Gestão. Manual de catalogação no Portal Brasileiro de Dados Abertos*. Brasília, DF. <http://wiki.dados.gov.br/GetFile.aspx?File=%2fManuais%2fManual-de-Cataloga%C3%A7%C3%A3o-v1.0.pdf>
10. Gobierno de Colombia (2016a). *Guía para el uso y aprovechamiento de Datos Abiertos en Colombia*. MinTIC. <https://herramientas.datos.gov.co/sites/default/files/Guia%20de%20Datos%20Abiertos%20de%20Colombia.pdf>
11. Gobierno de Colombia (2016b). *Guía para la estandarización de datos abiertos*. MinTIC. https://herramientas.datos.gov.co/sites/default/files/Guia_Estandarizacion_DatosAbiertos_final.pdf

12. Gobierno de Colombia (2016c). *Manual de datos abiertos en Colombia para publicador*. MinTIC. https://herramientas.datos.gov.co/sites/default/files/CO_417_MNL_PUBLICADOR_0.pdf
13. Gobierno de Colombia (2016d). *Manual de datos abiertos en Colombia para administrador*. MinTIC. https://herramientas.datos.gov.co/sites/default/files/CO_417_MNL_ADMINISTRADOR.pdf
14. Gobierno de Colombia (2021). ¿Qué son datos abiertos? <https://gobiernodigital.mintic.gov.co/portal/Iniciativas/Datos-abiertos/>
15. Gobierno de España (2012). Ministerio de Industria, Energía y Turismo, Ministerio de Hacienda y Administraciones Públicas. Guía de aplicación del Real Decreto 1495/2011. <http://datos.gob.es/es/documentacion/guia-de-aplicacion-del-real-decreto-14952011>
16. Gobierno de España (2015). Guía para publicar datos abiertos de manera rápida y sencilla (con CKAN). <http://datos.gob.es/es/documentacion/guia-para-publicar-datos-abiertos-de-manera-rapida-y-sencilla-con-ckan>
17. Gobierno de España (2016a). Agencia Española de Protección de Datos. Orientaciones sobre la protección de datos en la reutilización de la información del sector público. <http://datos.gob.es/es/documentacion/orientaciones-sobre-la-proteccion-de-datos-en-la-reutilizacion-de-la-informacion-del>
18. Gobierno de España (2016b). Agencia Española de Protección de Datos. Orientaciones y garantías en los procedimientos de anonimización de datos personales. <http://datos.gob.es/es/documentacion/orientaciones-y-garantias-en-los-procedimientos-de-anonimizacion-de-datos-personales>
19. Gobierno de España (2016c). Ministerio de Hacienda y Administraciones Públicas. Guía de aplicación de la Norma Técnica de Interoperabilidad. <http://datos.gob.es/es/documentacion/guia-de-aplicacion-de-la-norma-tecnica-de-interoperabilidad-de-reutilizacion-de>
20. Gobierno de España (2017a). Iniciativa Aporta. Manual de usuario: perfil organismo público. <http://datos.gob.es/es/documentacion/guias-de-datosgobes>
21. Gobierno de España (2017b). Iniciativa Aporta. Plantilla RDF/XML para la federación de conjuntos de datos. <http://datos.gob.es/es/documentacion/guias-de-datosgobes>
22. Gobierno de Portugal (2016). Dados.gov. <http://www.dados.gov.pt/pt/noticias/guia-dados-abertos.aspx#sthash.YbTEwKOP.fCIW0V2O.dpb>
23. Governo do Estado São Paulo. (2015). *Guia de dados abertos*. <http://www.governoaberto.sp.gov.br/wp-content/uploads/2016/02/Book-Web-Guia-de-Dados-Abertos.pdf>
24. Farias, Maria Giovanna; Farias, Gabriela; Sousa, Francisca; Abreu, Patrícia; Barreto, Rayar; Cabral, Viviane (2021). Profissionais da Informação: atuação resiliente durante a pandemia de Covid-19. *Brazilian Journal of Information Science*, 15(1), 17. <https://dialnet.unirioja.es/servlet/articulo?codigo=8040147>
25. França, Maira; Carvalho, Angela (2019). Análise da produção científica sobre mídias sociais e bibliotecas nos periódicos da Ciência da Informação no Brasil. *Em Questão*, 25(3), 296-320. <https://www.seer.ufrgs.br/EmQuestao/article/view/89958>
26. Gil, Antônio (2002). *Como elaborar projetos de pesquisa* (4.ª ed.). São Paulo: Atlas.
27. Gilliland, Anne (2008). Setting the stage. En Murtha Baca (Ed.), *Introduction to metadata* (2.ª ed.) (pp. 1-19). Los Angeles: Getty Research Institute,
28. Global Editors Network (2017). *Data Journalism Awards*. <https://www.datajournalismawards.org>
29. Le Coadic, Yves-François (2004). *A ciência da informação*. Brasília: Brinquet de Lemos.
30. Lucci, Elian; Branco, Anselmo; Mendoça, Cláudio (2013). *Território e sociedade no mundo globalizado* (2.ª ed.). São Paulo: Saraiva.
31. Mcgarry, Kevin (1999). *O contexto dinâmico da informação: uma análise introdutória*. Brasília: Briquet de Lemos.
32. Moreira, Fabio; Zafalon, Zaira; Santos, Plácida; Sant'Ana, Ricardo César Gonçalves (2017). Metadados para descrição de datasets e recursos informacionais do "Portal Brasileiro de Dados Abertos". *Perspectivas em Ciência da Informação*, 22(3), 158-185.
33. Moreira, Fábio; Valentim, Marta; Sant'Ana, Ricardo César Gonçalves (2018). A interdisciplinaridade da Ciência da Informação e suas contribuições no estudo do compartilhamento de dados governamentais na internet. *Em Questão*, 24(2), 300-329.
34. Organización para la Cooperación y el Desarrollo Económico (2017). *Government at a Glance 2017*. París: OECD Publishing. http://dx.doi.org/10.1787/gov_glance-2017-en

[Abordajes de metadatos para publicadores de datos gubernamentales en el contexto de Iberoamérica]

35. Organización para la Cooperación y el Desarrollo Económico (2018). *Members and partners*. <https://www.oecd.org/about/members-and-partners/>
36. Organización para la Cooperación y el Desarrollo Económico (2019). *Government at a Glance 2019*. París: OECD Publishing. <https://doi.org/10.1787/8ccf5c38-en>
37. Portal Europeo de Datos (2017). Preparing data. <https://www.europeandataportal.eu/pt/providing-data/goldbook/preparing-data>
38. Rosenfeld, Louis; Morville, Peter; Arango, Jorge (2015). *Information architecture: For the web and beyond*. Sebastopol: O'Really.
39. Sant'Ana, Ricardo (2013). Prefácio. En Rachel Alves; Plácida Santos (Eds.), *Metadados no domínio bibliográfico* (pp. 7-9.). Rio de Janeiro: Intertexto.
40. Santos, Plácida (2010). *Catálogo revisitado: sua história e contemporaneidade*. 248 f. (tese Livre-Docência). Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília.
41. Santos, Plácida; Simionato, Ana; Rodrigues, Fernando; Alencar, Cely; Ferreira, Jaider; Silva, Renata...; Lopes, Rita (2013). Mapeamento do termo tecnologia em periódicos da CI no escopo do GT8: informação e tecnologia. En *Encontro Nacional De Ciência Da Informação*, 14, Florianópolis. Anais... Florianópolis: ANCIB, 2013. <http://enancib.sites.ufsc.br/index.php/enancib2013/XIVenancib/paper/view/201>
42. Santos, Plácida; Sant'Ana, Ricardo César Gonçalves (2013). Dado e granularidade na perspectiva da informação e tecnologia: uma interpretação pela Ciência da Informação. *Ciência da Informação, Brasília*, 42(2), 199-209. <http://revista.ibict.br/ciinf/article/view/1382/1560>
43. Santos, Plácida; C., Simionato, Ana Carolina; Arakaki, Felipe (2014). Definição de metadados para recursos informacionais: apresentação da metodologia BEAM. *Informação & informação*, 19(1), 146-163. <http://hdl.handle.net/11449/114736>
44. Schema. (2017). Dataset. <http://schema.org/Dataset>
45. Silva, Silvado; Soares, Ana Thereza; Cesar, Daniel; Rabelo, Leon (2020). Indicadores para avaliação qualitativa de Dados Abertos: Inteligibilidade, operacionalidade e interatividade nos datasets do Governo Federal no Portal Brasileiro de Dados Abertos. *Informação & Sociedade: Estudos*, 30(3), 1-19. <https://doi.org/10.22478/ufpb.1809-4783.2020v30n3.52469>