



Palabra Clave (La Plata), Octubre 2018, vol. 8, n° 1, e056. ISSN 1853-9912  
Universidad Nacional de La Plata.  
Facultad de Humanidades y Ciencias de la Educación.  
Departamento de Bibliotecología

## Publicación de datos abiertos en instituciones de patrimonios culturales

---

### Débora Marroco Ninin

Universidade Federal de São Carlos (UFSCar), Brasil  
debora.mninin@gmail.com

### Ana Carolina Simionato

Universidade Federal de São Carlos (UFSCar), Brasil  
simionato.ac@gmail.com

**Cita sugerida:** Ninin, D. M. y Simionato, A. C. (2018). Publicación de datos abiertos en instituciones de patrimonios culturales. *Palabra Clave (La Plata)*, 8(1), e056. <https://doi.org/10.24215/18539912e056>

Recibido: 23 de mayo de 2018 - Aceptado: 22 de octubre de 2018 - Publicado: 31 de octubre de 2018



Esta obra está bajo licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional  
[http://creativecommons.org/licenses/by-nc-sa/4.0/deed.es\\_AR](http://creativecommons.org/licenses/by-nc-sa/4.0/deed.es_AR)

## Publicación de datos abiertos en instituciones de patrimonios culturales

Publication of open data in cultural heritage institutions

*Débora Marroco Ninin*

*Universidade Federal de São Carlos (UFSCar), Brasil*

debora.mninin@gmail.com

*Ana Carolina Simionato*

*Universidade Federal de São Carlos (UFSCar), Brasil*

simionato.ac@gmail.com

### RESUMEN:

A partir del creciente movimiento de datos abiertos y de los debates acerca de la publicación de datos, se cuestiona como los centros de información, pueden incitar a la publicación de estos datos en un formato abierto para facilitar su difusión, acceso, uso y reutilización. Por lo tanto, el objetivo es presentar una propuesta de publicación de datos abiertos para las instituciones del patrimonio cultural, elaborada en base a los resultados de un estudio comparativo entre los aspectos de metadatos relacionados con la descripción de patrimonios culturales y las buenas prácticas de publicación de datos recomendadas por el Consorcio World Wide Web (en adelante W3C). Los métodos y procedimientos incluyen la investigación exploratoria en la literatura científica y documental, nacional e internacional, acerca del movimiento de publicación de datos abiertos y sobre el uso de metadatos en la representación de patrimonios culturales. A partir de los resultados obtenidos, se concluye que existe una confluencia entre los objetivos del W3C, los procesos de representación practicados por la Ciencia de la Información y el alcance general de las Humanidades digitales, con respecto a la sostenibilidad de los datos publicados y de la gestión de los mismos en el ambiente digital. Se propone un procedimiento de publicación de datos abiertos para ser aplicado por las instituciones culturales, como un incentivo para el entendimiento y uso de las herramientas del movimiento por las instituciones culturales nacionales.

**PALABRAS CLAVE:** Publicación de datos, Metadatos, Patrimonio cultural.

### ABSTRACT:

From the growing movement of open data and discussions about data publishing, it is questioned: how can information centers incite the publication of these data in an open format to facilitate the dissemination, access, use and reuse? The objective is, then, to present the proposal for the publication of open data for cultural heritage institutions, from the results of a study between the aspects of metadata related to the description of assets and open data publishing practices that are recommended by the World Wide Web Consortium (W3C). The methods and procedures included exploratory research in the scientific and documentary literature, national and international, on the movement of data publishing and on the use of metadata in the representation of cultural heritage. The results showed a confluence between the objectives of the W3C consortium, the representation processes practiced by the information science and the general scope of the digital humanities with respect to the sustainability of the published data and the data management in the digital environment. It proposes a procedure for publishing open data to be applied by cultural institutions, as an incentive for the understanding and use of the tools of the movement by national cultural institutions.

**KEYWORDS:** Data publishing, Metadata, Cultural heritage.

## 1 INTRODUCCIÓN

La llegada y la difusión de las tecnologías de la información y comunicación (TIC) a todos los sectores de la sociedad implican cambios en los modos de producción, distribución y uso del conocimiento en cada uno de estos segmentos. De manera especial, destacamos las influencias observadas en las Ciencias Humanas y Sociales, que culminaron en la necesidad de pensar la inserción de las tecnologías en las actividades diarias de estas ciencias bajo la institución de un campo multidisciplinario, como ha estado sucediendo con las llamadas 'Humanidades Digitales'.

En este sentido, los centros de información y memoria -como los archivos, las bibliotecas, los museos y las galerías de arte -han buscado apropiarse de estas tecnologías utilizándolas en sus colecciones digitales a fin de alcanzar la integridad de estos objetivos. De esta manera, se destacan los temas relativos al acceso a la información cultural y a la manipulación de datos, ya que están directamente relacionadas con el creciente movimiento de la publicación de datos en diferentes dominios y, principalmente, se vinculan a todo lo referente a publicación de datos en formato abierto, que surgen del movimiento de datos abiertos a diferentes niveles de estructuración, pero de manera abierta.

Para una mejor aclaración del objeto de estudio, el patrimonio cultural es definido por Hyvönen como "[...] Legado de objetos físicos, ambientes, tradiciones y conocimientos de una sociedad, que pertenecen al pasado, que son conservados y desarrollados en el presente y que son preservados (conservados) en beneficio de las generaciones futuras" (2012, p. 1; nuestra traducción). En este enfoque, el patrimonio cultural puede dividirse en material e inmaterial. Ambientes, tradiciones y conocimientos integran el patrimonio cultural inmaterial, que puede estar registrado físicamente o no. En contraposición, el patrimonio cultural material, tiene una composición documental presente y apunta a las colecciones bibliográficas y a los archivos que registran el conocimiento humano (Dahlström, Hansson, y Kjellman, 2012; Maroevic, 1998; Souza y Crippa, 2010).

A partir de la composición documental, el patrimonio cultural es inherente a las tipologías abarcadas por el área de la Ciencia de la información. Además, lo relativo al acceso y el uso de la información están directamente relacionadas con el ámbito de la representación de la información y con el uso de metadatos para la descripción de los recursos informativos. Simionato, Arakaki y Costa Santos (2017) resumen que, hasta entonces, los procesos de representación en bibliotecas, archivos y museos buscaban satisfacer las necesidades específicas de la tipología de sus colecciones, de forma que la gestión de metadatos por medio de las tecnologías actuales elimina estas fronteras y proporciona modos comunes e interoperables de acceso, uso y reutilización de recursos.

Así, los procesos de representación de los patrimonios culturales deben armonizarse entre los procedimientos de bibliotecas, museos y archivos, además de buscar una apropiación de los nuevos procedimientos y herramientas tecnológicas frente a las propuestas de Humanidades Digitales. Por esta razón, se pregunta: ¿cómo los centros de información pueden incitar a la publicación de esos datos en un formato abierto para propiciar la difusión, acceso, uso y reutilización?

De ese modo, el objetivo del presente trabajo es presentar la propuesta para la publicación de datos abiertos en las instituciones de patrimonios culturales, procedentes de los resultados de un estudio comparativo entre los aspectos de los metadatos relacionados con la descripción de patrimonios culturales, y las buenas prácticas de publicación de datos recomendadas por el Consorcio World Wide Web (W3C).

Para describir los resultados, se utilizó el método perspectivista (Peterson, 1996). Con origen en la filosofía, el perspectivismo fue introducido en los estudios de la Ciencia de la Información por Costa Santos y Vidotti (2009). Así, este trabajo buscó aplicar el método perspectivista para conducir una investigación exploratoria en la literatura científica y documental, nacional e internacional, acerca del movimiento de publicación de datos y del uso de metadatos en la representación de patrimonios culturales.

## 2 PUBLICACIÓN DE DATOS

Puesto que son abordados los aspectos relacionados con los datos y los metadatos, es importante establecer los límites entre cada uno de estos conceptos dentro del ámbito de la ciencia de la información. Costa Santos y Santana (2013) definen los datos como la descomposición máxima de una información, constituido de una entidad (objeto o recurso), su atributo y el valor de este atributo. En el área de la representación de la información, estos atributos son las características necesarias para la descripción de los recursos de información, recibiendo la denominación de metadatos, definidos como atributos que representan una

entidad, y además aún “[...] son elementos descriptivos o atributos de referencia codificados que representan características propias o atribuidas a las entidades” (Alves, 2010, p. 47).

Pomerantz (2015) define que los metadatos son declaraciones sobre datos. Se considera, entonces, que el dato es constituido del metadato más el valor del recurso descrito. Por ejemplo, si el título de una película es *‘Barefoot in the park’* la obra en sí es la entidad/recurso, el atributo *‘title’* es el metadato, y *‘Barefoot in the park’* constituye el valor.

Los metadatos se toman como formas de comprensión básica para el proceso de estructuración de los datos, por esta razón, en documentos sobre las orientaciones de buenas prácticas para la publicación de datos, el uso de los metadatos es recomendado a menudo, y la mejor calidad de los metadatos en el momento de la descripción, asignando una utilidad significativa y acceso de aquel conjunto de datos, también denominado *datasets*.

Bizer, Heath, y Berners-Lee (2009) explican que cualquier persona puede publicar datos, que éstos pueden ser de cualquier tipo y que, por lo tanto, la web es genérica. Así, se pueden encontrar varios tipos de conjuntos de datos en la actualidad, publicados por individuos o por instituciones, con diversos objetivos y áreas de especialización. También cabe destacar que la principal característica para la publicación de datos abiertos en instituciones que tienen patrimonio cultural es que muchos de estos datos ya están estructurados de modo que sean comprensibles por una máquina, que surge de los registros de sus objetos en catálogos u otros tipos de herramientas de investigación.

Pensando en la coherencia y la homogeneidad en la publicación de datos, el W3C creó un grupo de trabajo para investigar las mejores prácticas de publicación de datos en la *Web*. Este grupo elaboró el documento "Mejores prácticas para publicar datos en la Web" agrupando un conjunto de aspectos para la publicación de datos (World Wide Web Consortium, 2017), destacando la necesidad de asignar metadatos descriptivos a los datos, siendo esta actividad típica de bibliotecarios, archiveros y museólogos en los procesos de representación de sus colecciones.

La propia literatura sobre el patrimonio cultural registra consideraciones acerca de las perspectivas relacionadas a la publicación de datos y de los actores involucrados en este proceso. Smith (2011) declara que la publicación de datos permite la creación de nuevos servicios, la construcción de alianzas entre instituciones que no necesariamente precisan pertenecer al mismo país y, principalmente, facilitan el desarrollo de conocimientos que sólo pueden ser alcanzados con datos abiertos y disponibles para la manipulación, el uso y la reutilización por diferentes actores de la sociedad. Aunque se centran en el patrimonio cultural, estas consideraciones pueden aplicarse en otros campos de la publicación de datos, especialmente en el ámbito científico.

Se observa que, en el ámbito de la publicación de los datos en la web, la unión de datos para que el conjunto sea abierto, no es la única exigencia. En lo que se refiere al movimiento de datos abiertos, consiste en publicar datos en la web bajo licencias abiertas (Arakaki, 2016) y no debe confundirse con el movimiento de datos abiertos enlazados, que incluye la aplicación de los principios de datos enlazados en datos abiertos (Berners-Lee, 2006).

Los datos abiertos, según la Open Knowledge (2009), deben contener las siguientes propiedades:

- Disponibilidad y acceso: la información debe estar disponible como un todo y a un costo razonable de reproducción, preferiblemente descargándola de internet. Además, la información debe estar disponible en una forma conveniente y modificable.
- Reutilización y redistribución: los datos deben ser provistos bajo términos que permitan reutilizarlos y redistribuirlos, e incluso integrarlos con otros conjuntos de datos.
- Participación universal: todos deben poder utilizar, reutilizar y redistribuir la información. No debe haber discriminación alguna en términos de esfuerzo, personas o grupos. Restricciones “no comerciales” que impiden el uso comercial de los datos; o restricciones de uso sólo para ciertos propósitos (por ejemplo, sólo para educación) no son permitidas.

Como se describe en la propia guía, hay muchos grupos de personas y organizaciones que pueden beneficiarse de la disponibilidad de datos abiertos. A partir de estos, sería más sensato tomar decisiones sobre procesos y productos, así como también, reflexionar sobre el potencial de la innovación en varios segmentos.

Hyvönen (2012) divide y lista las ventajas de publicar datos entre aquellos relacionados a los usuarios o consumidores, y a sus editores, vistos en la tabla 1.

Tabla 1 – Ventajas de editores y consumidores de datos

Editores	Consumidores
Creación de contenidos distribuidos	Visión global de los contenidos distribuidos y heterogéneos
Enriquecimiento semántico de los contenidos	Agregación automática de contenidos
Reutilización de los contenidos publicados	Posibilidad de búsquedas semánticas

Fuente: Adaptado de Hyvönen (2012)

La agregación automática de contenido y las búsquedas semánticas son proporcionadas debido a la publicación de datos de forma estructurada. Es importante destacar que, en la web, los editores y los consumidores son roles compartidos por los mismos actores, para que se pueda decir que tales ventajas se complementan entre sí y son interdependientes.

En el documento "Mejores prácticas para publicar datos en la Web" se detallan las directrices para los editores de datos, así como también para desarrollar la confianza en los datos y en las tecnologías entre los desarrolladores y otros actores sociales que se ocupan de la publicación de los datos que no dependen de las tecnologías utilizadas ni de los campos de actuación de los editores, así como un conjunto de 35 mejores prácticas a adoptar en la publicación de datos, distribuidas entre 13 aspectos:

- Metadatos: prácticas que señalan la importancia de los metadatos descriptivos y estructurales para la comprensión humana y el procesamiento automático por máquinas;
- Licencia: agregar metadatos de licencia de uso de determinado *dataset*;
- Procedencia: asignación de metadatos de origen del *dataset*,
- Calidad: asignar metadatos relacionados a la calidad de los datos;
- 'Versionado': prácticas de inclusión de metadatos relativos a la versión actual del dataset, así como la administración del histórico de versiones anteriores;
- Identificación: prácticas relacionadas con el uso de URIs;
- Formatos: conjunto de prácticas relativas a los múltiples formatos de datos legibles por máquina recomendados para cada tipo de dataset;
- Vocabulario: prácticas relativas al uso del vocabulario y al grado de formalidad apropiado para cada situación;
- Acceso: conjunto de prácticas que promuevan el acceso a los datos de forma rápida y actualizada;
- Preservación: prácticas relacionadas con el archivado de datos y de informaciones referente al dataset (vocabularios usados, recursos relacionados, etc.);
- Retroalimentación: proporciona opciones de retroalimentación de los consumidores de datos;
- Enriquecimiento: prácticas relacionadas con el enriquecimiento de datos y la disponibilidad de presentaciones complementarias;
- Reedición: prácticas a adoptar en caso de reedición de datos por otros editores (W3C, 2017).

Se observa que el primer aspecto, relacionado con el uso de metadatos para la descripción y estructuración de registros, se refleja en las otras prácticas indicadas por el consorcio, especialmente en los aspectos de la asignación de metadatos relativos a la licencia, procedencia, calidad y 'versionado'. En el contexto de la

publicación de datos, los metadatos descriptivos se refieren a un conjunto de metadatos requeridos para la descripción de un conjunto de datos tomado como un único recurso. Sin embargo, para los conjuntos de datos de los patrimonios culturales, también deben tenerse en cuenta los metadatos descriptivos para hacer la descripción de cada recurso que los compone.

Berners-Lee (2006) ha compilado el ranking de *5-star Linked Open Data* para la evaluación de los datos publicados. La vista de ranking se muestra en la figura 1.

Figura 1 – Ranking “5 estrellas” para publicación y conexión de datos



Fuente: Kim y Hausenblas, 2015.

En este ranking, se observa que los datos pueden estar abiertos, pero no necesariamente enlazados, ya que la aplicación de los principios de datos enlazados (*Linked Data*) se produce a partir de datos de ‘cuatro estrellas’, mediante el uso del *Uniform Resource Identifier* (URI) y del modelo *Resource Description Framework* (RDF). Es decir, la posibilidad de vincular los datos no es un requisito obligatorio en la publicación de datos, por lo que se entiende también que esta clasificación es acumulativa y que los datos abiertos pueden llegar a la clasificación “tres estrellas”, cuyos requisitos son la publicación bajo licencias abiertas, en formatos legibles por máquina y, al mismo tiempo, no propietarios.

Luego, a partir de las consideraciones de este documento sobre los metadatos descriptivos y su importancia como una buena práctica en la publicación de datos, se presentan cuestiones relacionadas a los metadatos en los procesos de descripción de patrimonios culturales, con enfoque en los aspectos que se alinean con el movimiento de publicación de datos.

### 3 METADATOS PARA LA DESCRIPCIÓN DEL PATRIMONIO CULTURAL

A partir de la definición de metadatos citada anteriormente, se cuestiona cuáles son los atributos básicos necesarios para describir los recursos que se reflejan en una representación real y efectiva.

Para Glushko describir “[...] es identificar y caracterizar la esencia, o la cuestión primordial, del recurso” (2014, p. 8). Gilliland (2016) aclara que los recursos tienen contenidos intrínsecos, contextos extrínsecos y estructuras que pueden tener ambas propiedades (intrínsecas o extrínsecas). En este sentido, el proceso de representación implica la descripción de los contenidos, así como los contextos y las estructuras que reflejan este contenido y las diferentes formas de acceso posibles al recurso en el ambiente en el que se está insertado.

Así se emplean, principalmente, las consideraciones de Gilliland (2016) y Riley (2017) sobre las características, las categorías y las funciones de los metadatos, aplicables según el contexto en que se utilicen.

La tabla 2 muestra los tipos o las categorías de metadatos, listados por Riley (2017), con ejemplos de propiedades y de uso para cada uno de ellos.

Tabla 2 – Tipología de los metadatos

Tipos de metadatos		Ejemplos de propiedades	Usos primarios
Descriptivos		Título Autor Asunto Género Fecha de publicación	Descubierta Presentación Interoperabilidad
Estructuras		Secuencia Posición en la jerarquía	Navegación
Lenguajes de marcado		Párrafo Cabecera Lista Nombre Fecha	Navegación Interoperabilidad
Administrativos	Técnicos	Tipo del archivo Formato del archivo Tiempo/Fecha de creación Esquema de compresión	Interoperabilidad Mantenimiento de objetos digitales Preservación
	Preservación	Verificación de los datos Acción de preservación	Interoperabilidad Mantenimiento de objetos digitales Preservación
	Derecho de autor	Status do copyright Términos de licencia Titulares de los derechos	Interoperabilidad Mantenimiento de objetos digitales

Fuente: Adaptado de Riley (2017)

Los metadatos descriptivos son los más comunes y difundidos en las prácticas de representación en la Ciencia de la Información. La tabla 2 también presenta metadatos técnicos, de preservación y de derechos de autor, agrupados como metadatos administrativos, ya que se ocupan del mantenimiento de los recursos y/o representaciones de estos con el tiempo (Riley, 2017). Finalmente, los metadatos estructurales manejan la organización del recurso en el ambiente en el que se encuentran, mientras que los lenguajes de marcado tratan directamente de la navegación y de la interoperabilidad entre sistemas de recuperación de información.

De esta manera, cada una de las categorías de metadatos se refleja en una función o propósito de uso. Gilliland (2016) lista como funciones de los metadatos: a) creación (en formatos únicos o múltiples), contextualización y reutilización de recursos, b) organización y descripción, c) validación de autenticidad y veracidad, d) búsqueda y recuperación, e) uso y preservación y f) indicación de disponibilidad.

Tradicionalmente, la Ciencia de la Información enfoca la organización, descripción, búsqueda y recuperación de recursos en bibliotecas, archivos y museos por medio de metadatos descriptivos, aunque también se utilizan otros tipos de metadatos (Pomerantz, 2015; Riley, 2017). La propia publicación de datos es un ejemplo de uso de otros tipos de metadatos.

En lo que se refiere a sus características, los metadatos poseen atributos relacionados con su fuente y modo de creación, naturaleza, status, estructura, semántica y nivel del recurso (Gilliland, 2016), y se aplican en el ambiente informacional de acuerdo con la formalidad necesaria (Costa Santos y Alves, 2009). Se destacan los atributos de naturaleza, semántica y estructura, ya que tratan directamente de aspectos relacionados con la descripción de los recursos en el campo de la Ciencia de la Información. Esto por qué Gilliland (2016) explica que, en relación a la naturaleza, los metadatos pueden ser creados por individuos sin conocimiento técnico específico o expertos y, en cuanto a la semántica, pueden tener contenidos controlados por vocabularios o se constituyen de lenguaje natural.

El atributo relativo a la estructura es el foco de la administración de metadatos, pues ellos pueden ser estructurados, semiestructurados o no estructurados (Almeida, 2002). El atributo 'estructura' no debe confundirse con el tipo estructural, ya que este atributo controla los metadatos en sí y no el recurso. La estructura, entonces, puede ser relativa al recurso, a la organización del recurso en el ambiente y al metadato que lo describe.

Alves (2005) distribuye los niveles de estructuración de metadatos en formatos simples, estructurados o ricos, considerando el formato simple: metadatos no estructurados, con semántica reducida y recuperación automática por robots; formatos estructurados: metadatos estructurados según normas emergentes; y, formatos ricos: metadatos complejos, con un alto grado de descripción por medio de normas, estándares y códigos específicos.

Las normas, los códigos y los estándares se aplican en todos los tipos de metadatos y, cuando se relacionan al atributo de semántica, en sus valores. A partir de la tipología elaborada por Gilliland (2016), en la tabla 3, se presentan ejemplos de los tipos de estándares de estructura, valores, contenidos e intercambio/formato de los datos, que pueden ser utilizados en el contexto de la representación de los patrimonios culturales:

Tabla 3. Estándares de metadatos para patrimonios culturales

Tipología	Estándares de metadatos
Estándares de estructura de datos	<i>Dublin Core Metadata Element Set</i> , <i>VRA Core Categories (VRA Core)</i>
Estándares de valores de datos	<i>LC Thesaurus for Graphic Materials (TGM)</i> , <i>Art &amp; Architecture Thesaurus (AAT)</i> , <i>Getty Thesaurus of Geographic Names (TGN)</i> , <i>The Union List of Artist Names (ULAN)</i> , <i>The cultural objects name authority (CONA)</i> , <i>ICONCLASS</i>
Estándares de contenido de datos	<i>Categories for the Description of Works of Art (CDWA)</i> , <i>Cataloging Cultural Objects (CCO)</i> , <i>SPECTRUM</i>
Estándares de intercambio y formato de datos	<i>CDWA Lite XML schema</i> , <i>Simple Dublin Core XML schema</i> , <i>Qualified Dublin Core XML schema</i> , <i>VRA Core 4.0 XML schema</i>

Fuente: Elaborado por los autores.

Los estándares de estructura normalizan los metadatos necesarios en la descripción de un determinado tipo de recurso. La comunidad de patrimonios culturales utiliza los estándares de estructura *Dublin Core Metadata Elements Set (DC)* y *VRA Core Categories (VRA Core)*.

El *Dublin Core* simple (DC) es un conjunto de quince metadatos básicos para la descripción de las características creadas por la *Iniciativa de Metadatos Dublin Core (DCMI)*. Forman parte de un conjunto mayor de metadatos padronizados y especificaciones técnicas para representación de recursos en la Web, denominado *DCMI Metadata Terms (DCTERMS)* o “Dublin Core Cualificado” (Dublin Core Cualificado) (Riley, 2017). Aunque no sea un estándar específico para patrimonios culturales, se utiliza a menudo en bibliotecas y repositorios digitales. El *VRA Core*, por otra parte, es un estándar propio para la descripción de trabajos de arte, cultura y sus imágenes digitales, desarrollado en colaboración entre la *Asociación de recursos visuales* y la *Biblioteca del Congreso* (Library of Congress, 2015).

Los estándares de valores y de contenidos no deben confundirse entre sí: los valores se refieren a los vocabularios controlados y a los términos utilizados por listas de autoridades, mientras que los estándares de contenido definen las reglas de sintaxis y de formato para relleno de los campos.

Entre los estándares que se muestran en la tabla 3, destacamos los elaborados por el *Getty Research Institute*, llamado vocabularios *Getty*. Su terminología cubre las artes en general, la arquitectura, las artes decorativas, los archivos, los materiales bibliográficos, y sus representaciones visuales, así como los aspectos de su conservación (Getty Research Institute, 2015). Integran los vocabularios *Getty* el *Art & Architecture Thesaurus (AAT)*, el *Getty Thesaurus of Geographic Names (TGN)*, la *The cultural objects name authority (CONA)* y *The Union List of Artist Names (ULAN)*.

Además de estos vocabularios, la tabla 3 también incluye el *LC Thesaurus for Graphic Materials (TGM)* y el *ICONCLASS*. El TGM es un vocabulario desarrollado por la Biblioteca del Congreso para la indexación de fotografías, impresiones, dibujos gráficos y otros materiales visuales por tema y género/formato (Library of Congress, 2017).



ICONCLASS (Waal,1972) fue creado en la década de los cincuenta por el profesor de Historia del Arte Henri van de Waal (1910-1972), de la Universidad de Leiden (Holanda) y consiste en un sistema de clasificación de asuntos para artes e iconografía, ampliamente utilizado por museos e instituciones de arte.

Todos estos vocabularios y listas de autoridades son utilizados en conjunto a los estándares de contenido *Categorías para la descripción de obras de arte* (CDWA), *Catalogación de objetos culturales* (CCO) y SPECTRUM.

El CDWA es un estándar desarrollado por el *Getty Research Institute* para la representación descriptiva de obras de arte, obras arquitectónicas y materiales culturales individuales o en grupos/colecciones, y que permite indexar imágenes de los recursos representados (Getty Research Institute, 2017). Se destaca que para Gilliland (2016) el CDWA es un estándar de estructura y no de contenido.

La *Catalogación de objetos culturales* (CCO) se constituye en un manual titulado "Guía para describir las obras culturales y sus imágenes" para el uso de la comunidad de patrimonios culturales, patrocinado por la *Fundación de Asociación de recursos visuales*, y que recomienda el uso de CDWA y vocabulario *Getty* (Visual Resources Association, 2006).

El SPECTRUM fue desarrollado por la *Asociación de Documentación del Museo* (MDA) (Day y Patel, 2002) para los procesos de gestión y documentación de los objetos museológicos, incluyendo su descripción y mantenimiento. Su última actualización, el Spectrum 5.0, está organizada en nueve procesos primarios, que incluyen los diferentes modos de adquisición de objetos (guardia, transferencia de título, permuta, entre otras), catalogación, reubicación y préstamo (para otros museos y/o exposiciones) (Collections Trust, 2017).

Finalmente, los estándares de formato e intercambio de datos utilizan ampliamente lenguajes de marcado para la integración e interoperabilidad entre sistemas (Riley, 2017). Los estándares indicados en la tabla 3 corresponden a codificaciones de los estándares de estructura, valores y contenidos en la lengua denominada *Lenguaje de marcado extensible* (XML). Este es un lenguaje de marcado recomendado por el consorcio W3C para el intercambio de datos, sea en Web o en otros ambientes digitales (World Wide Web Consortium, 2016) y su uso constituye un primer paso para la distribución de datos en la Web.

El uso de estos estándares promueve la representación de patrimonios culturales en diferentes niveles y para diferentes públicos. Los datos generados por estas descripciones pueden ser publicados para ampliar el acceso y uso de estas colecciones. Al respecto, a continuación, se presentan el debate acerca de las cuestiones relacionadas con el empleo de metadatos en el marco del movimiento de publicación de datos abiertos, el procedimiento de publicación desarrollado por Assumpção (2018) y una propuesta derivada de la publicación Datos abiertos que puede ser aplicada por instituciones culturales.

#### 4 DESCRIPCIÓN Y PUBLICACIÓN DE LOS DATOS DEL PATRIMONIO CULTURAL

Después de presentadas las consideraciones generales sobre la publicación de los datos abiertos y sobre las diferentes funciones, tipos y características de metadatos para la representación de recursos, se debaten cuáles de estos aspectos se relacionan entre los dos procesos, a partir del documento de las mejores prácticas publicadas por el consorcio W3C.

El consorcio aclara que la aplicación del conjunto de mejores prácticas, o principios de publicación, tiene como objetivo asegurar la reutilización, la comprensión, la posibilidad de conexión, el descubrimiento, la confianza, el acceso, la interoperabilidad y el procesamiento de los datos por consumidores y editores de datos, en diferentes situaciones y con visiones de largo plazo (World Wide Web Consortium, 2017). Al mismo tiempo, Hyvönen (2012) señala que la comunidad de patrimonios culturales debe considerar las cuestiones relacionadas con la autenticidad del contenido, los derechos de autor y la autorización para reproducir y utilizar los datos publicados. En otras palabras, esta comunidad tiene como enfoque la confianza, el acceso y la reutilización de los datos de patrimonios culturales.

En el mismo sentido, Gilliland (2016) afirma que los metadatos demuestran la autenticidad y el grado de integridad del recurso, establecen su contexto, identifican sus relaciones estructurales con otros recursos y proporcionan varios puntos de acceso para diferentes tipos de usuarios. De esta forma, se entiende que el uso sistemático de metadatos se alinea con los objetivos propuestos por el W3C (2017) con respecto a comprensión, descubrimiento, confianza y acceso a los datos de recursos así representados.

Es evidente la posibilidad de alineamiento entre los procesos de representación practicados por la Ciencia de la Información y las recomendaciones de las mejores prácticas del consorcio W3C, como se ha demostrado en la tabla 4. Esta presenta los tipos de metadatos listados por Riley (2017) y las características señaladas por Gilliland (2016) que se relacionan directamente con las mejores prácticas para la publicación de datos que implican el uso de metadatos y que se pueden aplicar en la descripción y la publicación de datos de patrimonios culturales.

Tabla 4. Tipos y características de metadatos para mejores prácticas de publicación de datos

Aspectos listados por el consorcio W3C	Tipos de metadatos	Características de metadatos
Metadatos descriptivos	Descriptivos	Semántica, estructura
Licencia	Administrativos (Derechos de autor)	
Procedencia	Descriptivos, Administrativos (Derechos de autor)	
Cualidad		Naturaleza, semántica y estructura
"Versionado"	Técnicos	
Formatos		Estructura
Vocabulario		Semántica
Acceso		Estructura
Preservación	Descriptivos, Administrativos (Preservación)	Semántica y Estructura

Fuente: Elaborado por los autores.

La descripción del recurso se destaca como una práctica importante en la publicación de datos y puede referirse tanto al conjunto de datos como a los datos que le componen. Posee como características aspectos semánticos y estructurales, expresados por estándares de estructura y de contenido desarrollados para diversos dominios, incluso por la comunidad de patrimonios culturales.

Los metadatos relativos a la licencia requieren una atención especial de esa comunidad, ya que deben considerar los derechos de autor de los recursos descritos como diferentes de los relacionados con el conjunto de datos. Así, por ejemplo, los derechos de una obra de arte son diferentes de los derechos sobre una imagen digital producida por la institución que representa esta obra. Los metadatos de licencia también están directamente relacionados con los de procedencia, ya que clarifican cuáles son las fuentes de datos y, por eso, promueven la confianza en los datos publicados.

La calidad de los datos está directamente relacionada con las decisiones tomadas por los editores (naturaleza laica o experta de los metadatos) en relación a los modos de estructuración y descripción (semántica) de los datos.

El 'versionado' tiene como objetivo preservar el conjunto de datos en relación con su histórico de desarrollo y actualización. Este aspecto de preservación no debe confundirse con la preservación de los datos o con los recursos descritos en el conjunto de datos. Estos están directamente relacionados a los metadatos descriptivos y de preservación que indican las condiciones físicas del recurso, sea un objeto tridimensional real o un objeto digital.

Prácticas relacionadas a los vocabularios se refieren a la estructuración semántica de los metadatos descriptivos del conjunto de datos, mediante el uso de ontologías y vocabulario de clases y propiedades.

Por último, los aspectos relacionados al acceso reanudan las características de estructuración de los datos y de los metadatos, pues estos deben ser legibles y recuperables por seres humanos y por máquinas. Esto demuestra la importancia de la administración de datos y metadatos frente al movimiento de publicación de datos. En este sentido, Lee, Allard, McGovern, y Bishop (2017) afirman que, en los ambientes de datos abiertos, los metadatos son esenciales para los procesos de curaduría digital.

A partir, entonces, de los objetivos del W3C para la elaboración del documento de mejores prácticas y de las consideraciones del propio documento, se puede afirmar que el Consorcio busca soluciones de sostenibilidad en la publicación de datos, con el objetivo de la disponibilidad y condiciones de acceso, uso y reutilización de datos a largo plazo. Estos objetivos convergen con los propuestos por las Humanidades Digitales, especialmente en lo que se refiere al acceso a la información cultural y a la posibilidad de manipulación de datos.

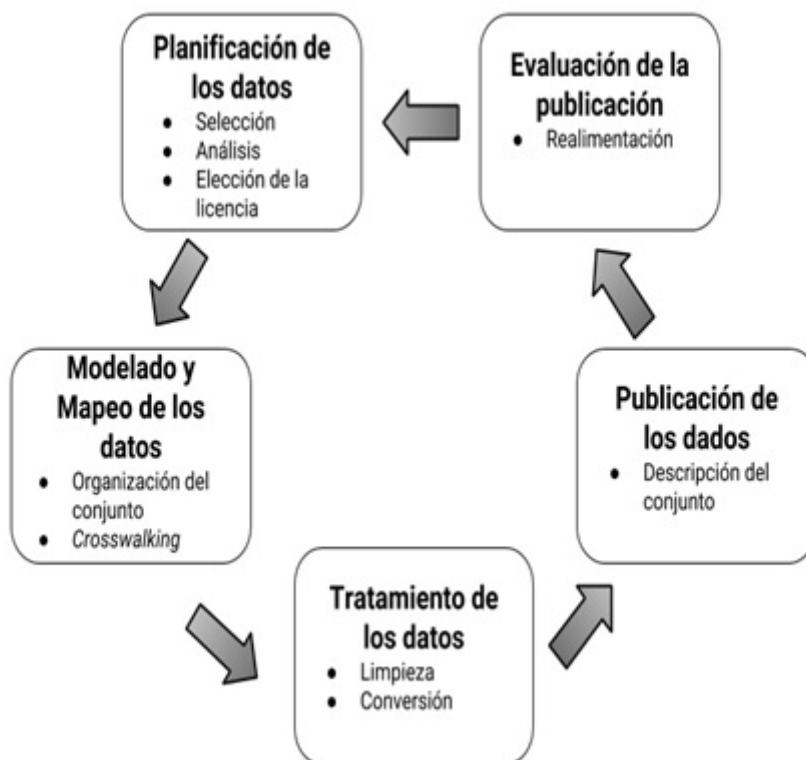
Maron, Yun, y Pickle (2013) explican que, en el contexto de las instituciones culturales, las cuestiones de sostenibilidad no se limitan a la preservación de los contenidos de sus colecciones para el futuro, pero abarcan su acceso, uso e impacto en la comunidad a largo plazo. Para el Council on Library and Information Resources (2001) la adopción de estándares y mejores prácticas, la construcción de arquitecturas digitales comunes y coherentes, así como la adopción de medios prácticos de creación y de intercambio de conocimientos son elementos primordiales para pensar la sostenibilidad en instituciones culturales.

La gestión de datos y metadatos en conjunto con el movimiento de publicación de datos abiertos actúan directamente sobre estos temas, ya que permiten el acceso, el uso y el intercambio de los contenidos digitales. Así, las bibliotecas, los archivos y los museos encuentran en la publicación de los datos una forma sostenible de gestionar sus colecciones y patrimonios culturales en el ambiente digital. Para lograrlo, es necesario pensar en rutinas y procedimientos para la publicación de datos abiertos.

Cabe destacar que ya se han realizado estudios para la publicación de datos de autoridad, como el trabajo de Assumpção (2018, p. 147) que elabora una propuesta de datos de autoridad y con el enfoque en publicaciones mediante el uso de los principios de datos enlazados. El modelo comprende los procesos de: 1) planeamiento, 2) modelado y mapeo, 3) tratamiento, relacionamiento y conversión, 4) publicación y 5) regeneración y retroalimentación de datos de autoridad. El proceso se inicia con la elección de los datos que se publicarán y las relaciones entre ellos, pensando en una publicación abierta vinculada a los datos. En el paso de modelado y mapeo se puede ver la inserción efectiva de los principios de datos enlazados, con la definición de los estándares para URIs para cada uno de los datos. Estos, cuando están publicados, deben ser accesibles a los seres humanos y a las máquinas (software) y promover la regeneración del conjunto de datos, con la inclusión de nuevos relacionamientos a partir del uso y reutilización de los datos.

Por esta razón, se propone el esquema para el *gerenciamiento de datos abiertos de objetos de patrimonio cultural* desarrollado en la figura 2 y adaptado de la propuesta de Assumpção (2018), que incluye las etapas de planificación, modelado y mapeo, tratamiento y conversión, y publicación de datos abiertos sobre colecciones de objetos de patrimonio cultural en bibliotecas, archivos y museos.

Figura 2. Gerenciamiento de datos abiertos de objetos de patrimonio cultural



Fuente: Elaborada por los autores.

La planificación comprende la selección de los datos que se publicarán, el análisis preliminar de dichos datos y la elección de licencias, que deben respetar el origen de los datos seleccionados, incluyendo la forma de recoger los datos de los objetos culturales. Assumpção (2018) dedica especial atención a los aspectos de la licencia y presenta en detalle la propuesta de licenciamiento *Creative Commons*, que puede ser utilizada por instituciones culturales en relación a sus conjuntos de datos.

El modelado y el mapeo engloban la organización lógica de los datos, con la realización de *cross walking* entre los estándares de metadatos utilizados en los casos que requieren la conversión de formatos de datos, como se indica en el siguiente paso, el de tratamiento y conversión. En este contexto, los datos se estructuran efectivamente en formatos abiertos, no propietarios y legibles por máquinas.

Para la publicación efectiva de los datos es necesario describir el conjunto de datos como un recurso único. En este sentido, para publicaciones de datos abiertos enlazados, Wood, et al. (2014) recomiendan el uso del *Vocabulary of Interlinked Datasets (VoID)* (Vocabulario de conjunto de datos interconectados), vocabulario en *RDF Schema* que utiliza el estándar *Dublin Core* para rellenar los campos de metadatos generales (autoría, título, licencia, descripción de datos), de acceso (HTTP URIs, por ejemplo) y estructurales del conjunto de datos. En base a estas recomendaciones para la publicación de datos abiertos, las instituciones culturales podrían, por ejemplo, asignar un HTTP URI y describir sus conjuntos de datos por medio del estándar *Dublin Core*, para facilitar su búsqueda y acceso.

La retroalimentación es un proceso ininterrumpido de actualización del conjunto de datos, con inclusión periódica de nuevos datos y verificación de consistencia de ellos. No aplicar los principios de datos enlazados dificulta los procesos de retroalimentación, ya que el mapeo de usos y reutilizaciones por parte de otros actores puede imposibilitar la producción de manera automática.

La simplificación de esta propuesta, en relación al propuesto por Assumpção (2018), sin la inclusión de los principios de datos enlazados, pretende fomentar las discusiones prácticas y las aplicaciones de publicación de datos abiertos por instituciones culturales, pues se percibe la dificultad de los profesionales de estas instituciones en dialogar y trabajar con los instrumentos involucrados en estos procesos.

Por lo tanto, con la aplicación de la propuesta presentada en la figura 2, instituciones culturales pueden iniciar procesos de publicación de datos que se mejoren y se amplíen con el tiempo, promoviendo un mayor acceso y uso de sus colecciones a partir de la inclusión de los principios de datos enlazados en sus publicaciones, de manera sostenible y adaptable a diferentes públicos.

## 5 CONSIDERACIONES FINALES

El movimiento de publicación de datos abiertos ofrece nuevas perspectivas acerca del acceso a las colecciones culturales, así como el uso y la reutilización de los datos publicados. A partir de los objetivos propuestos por la disciplina denominada Humanidades Digitales, se buscó comprender los beneficios que la publicación de datos ofrece para la difusión y el acceso al patrimonio cultural presente en las instituciones culturales, especialmente en bibliotecas, archivos y museos.

Del estudio comparativo entre las consideraciones de la literatura acerca de los tipos, las funciones y las características de los metadatos para la descripción de los recursos y sobre el uso de metadatos en la publicación de datos, expresado en el documento del W3C, se concluye que existe una convergencia entre los objetivos del consorcio, los procesos de representación practicados por la Ciencia de la Información y el alcance general de las Humanidades Digitales con respecto a la comprensión, el descubrimiento, la confiabilidad, la accesibilidad y la reutilización de datos publicados a largo plazo. Al mismo tiempo, demostraron que estos movimientos buscan soluciones de sostenibilidad en la gestión de datos en el ambiente digital.

A partir de estos resultados, fueron presentados y debatidos los procedimientos de publicación de datos de autoridad propuestos por Assumpção (2018) y su adaptación a publicaciones de datos abiertos de los patrimonios culturales. La adaptación se debe a la dificultad percibida en la aplicación de las herramientas del movimiento de datos abiertos enlazados por las instituciones culturales nacionales.

En vista de este escenario, existen perspectivas de estudios más específicos sobre cada una de las mejores prácticas de publicación, especialmente en lo que se refiere a los derechos de autor y a la asignación de licencias para el uso de datos relativos a patrimonios culturales. También se recomiendan otros estudios para revisar y ampliar el procedimiento propuesto a partir del trabajo de Assumpção (2018), cuyas limitaciones, en particular el aspecto de retroalimentación de los conjuntos de datos publicados, ofrecen barreras a los posibles usos de estos datos.

Se espera que este trabajo fomente el entendimiento y uso de las herramientas del movimiento de datos abiertos por las instituciones culturales nacionales, sirviendo como base para futuros proyectos mayores y más complejos de publicación de datos abiertos enlazados de datos de patrimonios culturales. Con las humanidades cada vez más digitales, las instituciones culturales encuentran en estas tecnologías formas de ampliar sus fronteras y llegar a nuevos públicos.

## REFERENCIAS

- Almeida, M.B. (2002). Uma Introdução Ao XML, Sua Utilização Na Internet e Alguns Conceitos Complementares. *Ciência Da Informação*, 31(2), 5–13. <https://doi.org/10.1590/S0100-19652002000200001>
- Alves, R.C.V. (2005). *Web Semântica: Uma Análise Focada No Uso de Metadados*. (Tesis de maestría). Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília (SP). Recuperada de <http://hdl.handle.net/11449/93690>

- Alves, R.C.V. (2010). *Metadados Como Elementos Do Processo de Catalogação* (Tesis doctoral). Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília (SP). Recuperada de <http://repositorio.unesp.br/handle/11449/103361>
- Arakaki, F.A. (2016). *Linked Data: Ligação de Dados Bibliográficos* (Tesis de maestría). Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília (SP). Recuperada de <https://repositorio.unesp.br/handle/11449/147979>
- Assumpção, F.S. (2018). *Modelo Para a Publicação de Dados de Autoridade Como Linked Data* (Tesis doctoral). Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília (SP). Recuperada de <http://hdl.handle.net/11449/152759>
- Berners-Lee, T. (2006). *Linked Data: Design Issues*. Recuperado de <https://www.w3.org/DesignIssues/LinkedData.html>
- Bizer, C., Heath, T., y Berners-Lee, T. (2009). Linked Data: The Story so Far. En A.P. Sheth, *Semantic Services, Interoperability and Web Applications: Emerging Concepts* (pp. 205–227). Estados Unidos: IGI Global.
- Collections Trust. (2017). *Introduction to Spectrum 5.0*. Recuperado de <https://collectionstrust.org.uk/spectrum/spectrum-5/>
- Costa Santos, P.L.V.A., y Alves, R.C.V. (2009). Metadados e Web Semântica Para Estruturação Da Web 2.0 e Web 3.0. *DataGramaZero: Revista de Ciência Da Informação*, 10(6). Recuperado de <http://www.brapci.inf.br/index.php/article/download/52958>
- Costa Santos, P.L.V.A., y Santana, R.C.G. (2013). Dado e Granularidade Na Perspectiva Da Informação e Tecnologia: Uma Interpretação Pela Ciência Da Informação. *Ciência Da Informação*, 42(2), 199–209. Recuperado de <http://dx.doi.org/10.18225/ci.inf.v42i2.1382.g1560>
- Costa Santos, P.L.V.A., y Vidotti, S.A.B.G. (2009). Perspectivismo e Tecnologias de Informação e Comunicação: Acréscimos à Ciência Da Informação? *DataGramaZero: Revista de Ciência Da Informação*, 10(3). Recuperado de [http://dgz.org.br/jun09/Art\\_02.htm](http://dgz.org.br/jun09/Art_02.htm)
- Council on Library and Information Resources. (2001). *Building and Sustaining Digital Collections: Models for Libraries and Museums*. Washington, D.C: Council on Library and Information Resources.
- Dahlström, M., Hansson, J., y Kjellman, U. (2012). As We May Digitize. *LIBER Quarterly* 21(3-4), 455-474. <https://doi.org/10.18352/lq.8036>
- Day, M., y Patel, M. (2002). *Metadata for Images: Report for the FILTER Project*. Recuperado de <http://www.ukoln.ac.uk/metadata/filter/report/report.html#3.7>
- Getty Research Institute. (2015). *Getty Vocabularies*. Recuperado de <http://www.getty.edu/research/tools/vocabularies/>
- Getty Research Institute. (2017). *Categories for the Description of Works of Art (CDWA)*. Recuperado de [http://www.getty.edu/research/publications/electronic\\_publications/cdwa/introduction.html#general](http://www.getty.edu/research/publications/electronic_publications/cdwa/introduction.html#general)
- Gilliland, A.J. (2016). Setting the Stage. En M. Baca (ed.) *Introduction to Metadata* (3ra. ed.) Los Angeles: Getty Research Institute. Recuperado de <http://www.getty.edu/publications/intrometadata/>
- Glushko, R.J. (Ed.). (2014). *The Discipline of Organizing: Core Concepts Edition*. Sebastopol, EUA: O'Reilly Media. Recuperado de <https://www.overdrive.com/search?q=70B2FB07-B1BE-458C-8BC5-5E7FBBEF946E>
- Hyvönen, E. (2012). Publishing and Using Cultural Heritage Linked Data on the Semantic Web. *Synthesis Lectures on the Semantic Web: Theory and Technology*, 2(1), 1–159. <https://doi.org/10.2200/S00452ED1V01Y201210WBE003>
- Kim, J.G., y Hausenblas, M. (2015). *Five Star Open Data*. Recuperado de <http://5stardata.info/en/>
- Lee, C., Allard, S., McGovern, N., y Bishop, A. (2017). Open Data Meets Digital Curation: An Investigation of Practices and Needs. *International Journal of Digital Curation*, 11(2), 115–25. <https://www.doi.org/10.2218/ijdc.v11i2.403>
- Library of Congress (2015). *Standards at the Library of Congress*. Recuperado de <https://www.loc.gov/librarians/standards>

- Library of Congress (2017). *Controlled Vocabularies*. Recuperado de <https://www.loc.gov/librarians/controlled-vocabularies/>
- Maroevic, I. (1998). The Phenomenon of Cultural Heritage and the Definition of a Unit of Material. *Nordisk Museologi*, 2, 135-142. Recuperado de <https://www.journals.uio.no/index.php/museolog/article/view/3829>
- Maron, N.L., Yun, J. y Pickle, S. (2013). *Sustaining Our Digital Future: Institutional Strategies for Digital Content*. Ithaca S+ R.
- Open Knowledge. (2009). *O Que São Dados Abertos?*. Recuperado de [http://opendatahandbook.org/guide/pt\\_BR/what-is-open-data/](http://opendatahandbook.org/guide/pt_BR/what-is-open-data/)
- Peterson, D. (1996). *Forms of Representation: An Interdisciplinary Theme for Cognitive Science*. Intellect Books.
- Pomerantz, J. (2015). *Metadata*. Cambridge, Massachusetts; London, England: The MIT Press.
- Riley, J. (2017). *Understanding Metadata: What Is Metadata, and What Is It For?*. Baltimore, MD: National Information Standards Organization (NISO). Recuperado de [http://www.niso.org/apps/group\\_public/download.php/17446/Understanding%20Metadata.pdf](http://www.niso.org/apps/group_public/download.php/17446/Understanding%20Metadata.pdf)
- Simionato, A.C., Arakaki, F.A., y Costa Santos, P.L.V.A. (2017). Descrição em bibliotecas, arquivos, museus e galerias de arte: Linkando recursos e comunidades. *Informação & Informação*, 22(2), 449-466. <http://dx.doi.org/10.5433/1981-8920.2017v22n2p449>
- Smith, M. (2011). *Proposed a 4-Star Classification-Scheme for Linked Open Cultural Metadata*. Recuperado de <http://lodlam.net/2011/06/06/proposed-a-4-star-classification-scheme-for-linked-open-cultural-metadata/>
- Souza, W.E.R., y Crippa, G. (2010). O campo da ciência da informação e o patrimônio cultural: Reflexões iniciais para novas discussões sobre os limites da área. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência Da Informação*, 15(29), 1-23. <https://doi.org/10.5007/1518-2924.2010v15n29p1>
- Visual Resources Association. (2006). *Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Images (CCO)*. Recuperado de <http://cco.vrafoundation.org/index.php/aboutindex/>
- Waal, H.v. (1972). *Iconclass: A Multilingual Classification System for Cultural Content*. Recuperado de <http://www.iconclass.nl/home>
- Wood, D., Zaidman, M., Ruth, L., y Hausenblas, M. (2014). *Linked Data: Structured Data on the Web*. Shelter Island, NY: Manning.
- World Wide Web Consortium. (2016). *Extensible Markup Language (XML)*. Recuperado de <https://www.w3.org/XML/>
- World Wide Web Consortium. (2017). *Data on the Web Best Practices*. Recuperado de <https://www.w3.org/TR/dwbp/>