

Análisis de tipos de descripción de fotografía antigua en Wikimedia Commons

Jesús Tramullas¹, Ana I. Sánchez-Casabón² y Piedad Garrido-Picazo³

¹ ORCID [0000-0002-5374-9993](https://orcid.org/0000-0002-5374-9993). Depto. Ciencias de la Documentación, Univ. de Zaragoza, España.

tramullas@unizar.es

² ORCID [0000-0002-0908-1615](https://orcid.org/0000-0002-0908-1615). Depto. Ciencias de la Documentación, Univ. de Zaragoza, España.

asanchez@unizar.es

³ ORCID [0000-0002-1750-7225](https://orcid.org/0000-0002-1750-7225). Depto. Informática e Ingeniería de Sistemas, Univ. de Zaragoza, España.

piedad@unizar.es

Resumen. Este trabajo revisa, usando una aproximación descriptiva, las características de análisis y descripción de los fondos de fotografía antigua disponibles en Wikimedia Commons. A tal fin, se han seleccionado seis fondos, proporcionados por instituciones GLAM. Sobre una selección de imágenes de estos fondos, se revisa si han sido descritos siguiendo los esquemas de metadatos Dublin Core, ISAD(G) o Europeana Data Model. Los resultados obtenidos permiten afirmar que no se aplican estándares de análisis documental ni esquemas de metadatos. El desarrollo de etiquetado semántico en Wikimedia Commons podrá ser una aproximación que ofrezca una solución a estas carencias.

Palabras clave: fotografía antigua, Wikimedia Commons, descripción, GLAM

Abstract. This paper reviews, using a descriptive approach, the characteristics of analysis and description of the ancient photography available on Wikimedia Commons. Six collections have been selected, provided by GLAM institutions. On a selection of 60 images, it is checked if they have been described following the Dublin Core, ISAD (G) or Europeana Data Model metadata schemas. The results obtained allow to affirm that description standards and metadata schemas are not applied. The development of semantic labeling in Wikimedia Commons may be an approach that offers a solution to these deficiencies.

Keywords: ancient photography, Wikimedia Commons, description, GLAM

1 Planteamiento

Wikimedia Commons se ha convertido en un importante repositorio para los fondos de fotografía, y en particular los de fotografía antigua, merced al desarrollo de proyectos de liberación de material digitalizado por parte de instituciones de la memoria (Tramullas y Ojeda, 2018). Estas instituciones participan a través del desarrollo de proyectos GLAM (*Galleries, Libraries, Archives & Museums*) los cuales se asocian a la liberación de materiales textuales, gráficos o sonoros, puestos a disposición de todos los usuarios, bajo licencias libres (Szajewski, 2013). Estas actuaciones se imbrican en el movimiento Open Access, que comenzó siendo una iniciativa de acceso abierto a las publicaciones científicas, pero que ha ido ampliando sus objetivos, hasta incluir elementos y colecciones de patrimonio cultural, que dadas las características del acceso, son resultantes de procesos de digitalización (Phillips, 2013). Además, se ha establecido una relación entre la presencia en Wikipedia y Commons de bibliotecas, museos y archivos, y de sus colecciones, y el aumento de consumo cultural, lo que resulta de gran importancia para facilitar el acceso, conocimiento y disfrute del patrimonio cultural (Navarrete y Borowiecki, 2016).

Wikimedia Commons sigue un proceso de aportación y edición colaborativa en abierto, al igual que Wikipedia. En el proceso colaborativo pueden participar tanto usuarios interesados, sin formación específica, como profesionales de la gestión de documentos. Este entorno abierto contrasta con los procesos de tratamiento de información que se llevan a cabo en las instituciones de la memoria, llevados a cabo por profesionales, y en el marco de un contexto normativo bastante estricto. En consecuencia, esta dicotomía lleva a plantear diferentes cuestiones sobre el nivel de detalle descriptivo usado en Commons, frente al aplicado en los catálogos especializados de las instituciones de la memoria. En primer lugar, hay que revisar el proceso de carga de imágenes y el esquema de descripción de contenido que se aplica, y sus características. En segundo lugar, la interacción que se lleva a cabo sobre la descripción, en especial los procesos de corrección o mejora de la misma.

Ambas cuestiones se trasladan a dos preguntas de investigación. La primera pregunta que cabe plantear es si estas prácticas se trasladan a Wikimedia Commons, o si se adoptan otros modelos descriptivos, influenciados por o adaptados al medio que los acoge. La segunda pregunta que se plantea es si las descripciones realizadas originalmente son aceptadas o son modificadas por la comunidad de editores de Commons, lo que abre un campo de investigación de interés sobre los procesos colaborativos de descripción y clasificación de documentos de imagen.

2 Metodología

La investigación sobre diferentes aspectos de Wikipedia y el resto de proyectos Wikimedia ha crecido notablemente en la última década (Tramullas, 2015). Este trabajo toma como punto de partida la propuesta clásica de atender a aspectos de

análisis formal. Dadas las características de los fondos de fotografía antigua disponibles en Commons, se ha optado por un análisis descriptivo de las características que presentan. En consecuencia, se revisará si los datos disponibles para cada fotografía responden a la información mínima requerida para una adecuada descripción del objeto. Aunque se ha señalado la dificultad que supone el abordar los procesos de análisis documental del objeto fotográfico, Sánchez-Vigil (2006: 180-181) propone la utilización de ISAD(G) para la descripción de documentos fotográficos. Iglésias-Franch (2008: 81-100) llevó a cabo una revisión de los diferentes estándares de metadatos para descripción y gestión existentes para fotografía digital, recomendando el uso de ISAD(G) etiquetado en EAD, aunque llamaba la atención sobre las limitaciones de esta aproximación.

Para el análisis de las estructuras de descripción y etiquetado de los fondos fotográficos se ha seguido la siguiente secuencia:

1. Identificación y revisión del protocolo de aportación de material fotográfico a Wikimedia Commons.
2. Identificación de los esquemas de descripción y categorización de material fotográfico.
3. Selección de fondos a revisar. A tal fin, se han identificado fondos específicos e imágenes aportados por seis instituciones de la memoria, en el marco de proyectos GLAM institucionales.
4. Revisión de los metadatos EXIF identificados.
5. Análisis de los esquemas de descripción utilizados. Comparación con las propuestas de Dublin Core (DC), ISAD(G) y Europea Data Model (EDM).
6. Análisis de los principios y esquemas de categorización utilizados para la descripción de las imágenes.
7. Revisión del historial de cambios de las imágenes en Wikimedia Commons.

3 Análisis de fondos fotográficos

El análisis que se expone en los siguientes párrafos ha sido realizado durante el mes de mayo de 2019. En entornos cuyos servicios, funcionalidades y contenidos evolucionan a lo largo del tiempo, la revisión de los resultados obtenidos puede ofrecer variaciones dependiendo de diferentes factores, como el período transcurrido o la incorporación o modificación de funcionalidades técnicas.

El proceso de carga y la descripción inicial de material en Wikimedia Commons ha sido detallado por Saorín y Pastor (2011) y por Tramullas y Ojeda (2018). El proceso básico no se presta a realizar una descripción detallada: basta con dar un nombre para el fichero del objeto, una descripción textual no estructurada, y asignar una o varias categorías, tomadas de la estructuras preexistentes, o de nueva creación. Commons puede recomendar categorías por similitud con el texto introducido por el

editor a tal fin. Las herramientas que permiten el proceso de carga por lotes tampoco proveen la capacidad de incorporar metadatos estructurados.

Como consecuencia de las características del proceso de carga, no se han identificado esquemas de metadatos preparados previamente para la descripción de las imágenes. En aquellos casos en que las imágenes incorporan elementos específicos, éstos han sido incorporados a las mismas en un proceso de edición y modificación de contenidos posterior. En varios casos estos elementos se han organizado utilizando plantillas creadas específicamente a tal fin por las instituciones responsables del proyecto. El uso de las categorías como esquema para clasificar las imágenes muestra una tendencia clara a incorporar descriptores que intentan desempeñar funciones de identificación del fondo de origen, así como de identificación de algunos elementos básicos, como la autoría o la cronología, pero sin que esto sea consistente en todas las fotografías y fondos revisados. El sistema de categorías de Wikipedia ha sido ampliamente utilizado en diferentes ámbitos de investigación (Tramullas, Sánchez-Casabón y Garrido-Picazo, 2018).

Para el análisis de los esquemas de descripción utilizados, se han seleccionado fotografías antiguas provenientes de seis proyectos GLAM llevados a cabo por instituciones, de manera oficial, en Wikimedia Commons. Los seis proyectos se encuentran recogidos en la tabla 1.

Tabla 1. Proyectos GLAM objeto de revisión.

| Institución | URL de referencia |
|---|---|
| <i>Library of Congress</i> | https://commons.wikimedia.org/wiki/Category:Collections_in_the_Library_of_Congress |
| <i>Deutsches Fotothek</i> | https://commons.wikimedia.org/wiki/Category:Images_from_the_Deutsche_Fotothek |
| <i>Metropolitan Museum of Art</i> | https://commons.wikimedia.org/wiki/Category:Images_from_Metropolitan_Museum_of_Art |
| <i>Georges Eastman House Collection</i> | https://commons.wikimedia.org/wiki/Category:Photographs_in_the_George_Eastman_House_Collection |
| <i>The J. Paul Getty Museum</i> | https://commons.wikimedia.org/wiki/Category:Google_Art_Project_works_in_The_J._Paul_Getty_Museum |
| <i>Fundación Joaquín Díaz</i> | https://commons.wikimedia.org/wiki/Category:Images_provided_by_Fundaci%C3%B3n_Joaqu%C3%ADn_D%C3%ADaz |

Para cada uno de los seis proyectos se han buscado y seleccionado diez fotografías. En cada una de ellas se han comprobado los metadatos EXIF existentes, y la descripción, la categorización, y el uso de esquemas de metadatos para la descripción del objeto fotográfico. Los resultados obtenidos se resumen en la tabla 2.

Tabla 2. Esquemas de metadatos identificados.

| Institución | EXIF | Metadatos | DC | ISAD(G) | EDM |
|---|---------|-----------|----|---------|-----|
| <i>Library of Congress</i> | Parcial | Si | No | No | No |
| <i>Deustches Fotothek</i> | Parcial | Si | No | No | No |
| <i>Metropolitan Museum of Art</i> | Parcial | Si | No | No | No |
| <i>Georges Eastman House Collection</i> | Parcial | Si | No | No | No |
| <i>The J. Paul Getty Museum</i> | Parcial | Si | No | No | No |
| <i>Fundación Joaquín Díaz</i> | Parcial | Si | No | No | No |

La aparición y formulación de los metadatos EXIF no es consistente. Dado que las imágenes revisadas provienen de diferentes políticas y procesos de digitalización, la presencia de los mismos es parcial, debido a las diferentes cabeceras generadas por las herramientas informáticas usadas en el proceso. Incluso en ocasiones se carece por completo de estos metadatos, que no han sido considerados relevantes por la institución generadora.

Figura 1. Ejemplo de descripción básica de fotografía antigua.

La columna etiquetada “Metadatos” corresponde a aquellas fotografías en las cuales se han añadido metadatos básicos descriptivos, del tipo autoría, cronología, título, ubicación, técnica... (véanse figuras 1 y 2). Sin embargo, y aunque puede establecerse una correspondencia con algunos de los elementos de otros esquemas de metadatos, como puede ser, por ejemplo, Dublin Core, realmente no se ha identificado en ninguna de las fotografías revisadas el uso ni de Dublin Core, ni de ISAD(G), ni de EDM.

| Rue de l'Hotel de Ville | |
|-----------------------------|---|
| Artista | Eugène Atget (1857-1927)    |
| Título | Rue de l'Hotel de Ville |
| Descripción | English: Eugène Atget's penchant for photographing in the early morning accounts for the street's empty stillness. A sole inhabitant peers out from the hotel doorway, so unobtrusive he becomes a part of the architecture as he quietly watches the photographer work. Atget's decision to photograph when the city was deserted heightens the sense of desolate abandonment in the city's older sections, soon to be demolished to make way for grand, broad, new boulevards. The two dark, solemn buildings in the foreground thus frame the old, narrow lane with the solid weight of history. An hourglass-shaped shaft of bright daylight cuts through the frame, dividing the street into shadow and illumination. |
| Fecha | 1921 |
| Técnica | Albumen silver print |
| Dimensiones | Image: 21.6 x 18.3 cm (8 1/2 x 7 3/16 in.) |
| Colección | Getty Center    |
| Número de inventario | 84.XM.1034.7 |
| Inscripciones | Title and negative number 6267 inscribed verso print in pencil. |
| Fuente/fotógrafo | The Getty Center, Object 62231  |
| |  <p>Esta imagen proviene del programa Open Content Program® (programa de contenido abierto) del Instituto de Getty Research, lo cual ha dicho que después de su evaluación, no hay conocidas restricciones del derecho de autor.</p> <p>Imágenes de Open Content son reproducciones digitales de obras de arte que están en las colecciones de Getty y en el dominio público. Tenemos los derechos de autor de estas imágenes, o no somos conscientes de cualquier restricción del uso de ellas.</p> <p>Mientras el Instituto de Getty Research no puede comentar con certidud al estado de derechos de una imagen en particular, ha dicho que se pueden utilizar las imágenes de Open Content por cualquier propósito sin la necesidad de permiso previo de Getty.</p> <p>Para más información, véase http://www.getty.edu/about/opencontent.html.</p> |

Figura 2. Ejemplo de descripción básica de fotografía antigua.

Cabe destacar que en la interfaz de uso de los documentos se muestra una pestaña titulada “Datos estructurados”, y cuyo contenido deberá ser provisto a través de la introducción de descripciones estructuradas de datos a través del proyecto *Commons:Structured data*. Sin embargo, este proyecto todavía se encuentra en desarrollo, y por el momento ofrece como esquema de metadatos una versión del provisto por CIDOC para objetos museísticos. La tabla de propiedades del proyecto tienen previsto ofrecer la descripción del objeto y de su contenido a través de la propiedad P180, “representa a”, y de los calificadores correspondientes. En el conjunto estudiado no se ha encontrado ninguna imagen que implemente esta propiedad. La utilización del sistema de categorías muestra una tendencia a intentar su uso como reflejo del esquema de organización de propio fondo, estableciendo la proveniencia y las series de origen (véase fig. 3). Sin embargo, y al igual que en otros aspectos, la aproximación usada tampoco es homogénea entre instituciones, con diferencias que pueden apreciarse en el uso de categorías como sustitutas de metadatos.

Categorías (+): 1921 photographs of Paris (-) (±) (↓) (↑) | 1921 photographs (-) (±) (↓) (↑)
 Photographs of Paris by Eugène Atget (-) (±) (↓) (↑)
 Photographs in The J. Paul Getty Museum (-) (±) (↓) (↑) | Rue de l'Hôtel-de-Ville (Paris) (-) (±) (↓) (↑)
 Images of Paris from the Getty Museum (-) (±) (↓) (↑) | (+)
 Categorías ocultas: Files from the Getty's Open Content Program | Artworks with known accession number
 Artworks without Wikidata item | CC-PD-Mark | PD-old-80

Figura 3. Ejemplo de uso de categorías como estructura del fondo.

Finalmente, la revisión de los historiales de edición recoge el ciclo de vida de los procesos de carga y modificación de las imágenes. La primera versión de las imágenes digitalizadas es la que prevalece, habiendo detectado sólo un cambio de fichero entre los 60 analizados. El número medio de ediciones que modifican la aportación y datos originales es de 7, la mayoría correspondientes a correcciones automáticas realizadas por un bot, o a correcciones de asignación de categorías, mientras que aquellas que añaden o modifican datos metadatos se reducen a un 18% del total.

4 Conclusiones

La descripción de los elementos u objetos fotográficos disponibles en Wikimedia Commons es irregular y carece de homogeneidad, ya que no existe un estándar predefinido que haya sido establecido acorde a los requerimientos establecidos para el tratamiento documental y archivístico de este tipo de materiales. La información aportada por los metadatos EXIF es mínima, ya que en aquellos casos en los que aparece sólo informa sobre la herramienta de digitalización utilizada. En lo que concierne a los aspectos documentales y archivísticos, no se han identificado, en los materiales seleccionados, esquemas de descripción que sigan lo establecido por DC, ISAD(G) o EDM. Tampoco ha podido establecerse la aplicación de esquemas uniformes de categorización o clasificación. El análisis de los diferentes grupos y elementos individuales revisados demuestra que diferentes instituciones han utilizado diferentes aproximaciones para la descripción y categorización de los fondos fotográficos.

La revisión de los historiales de actividad sobre las imágenes revela que en muy pocos casos son modificadas las descripciones y clasificaciones llevadas a cabo en el momento de la aportación a Commons, lo que induce a pensar que la comunidad de editores da por válida la información provista por las instituciones.

El desarrollo y puesta en marcha del proyecto *Commons:Structured Data*, que pretende ofrecer un marco semántico de descripción de objetos culturales utilizando la base de Wikidata (Vrandečić y Krötzsch, 2014)) se orienta a resolver las carencias de descripción existentes, y a facilitar la reutilización y enlazado del material disponible en otros recursos de información en productos de información abiertos (de Boer et alii, 2019).

Referencias

- de Boer, V., de Bruyn, T., Brooks, J., & de Vos, J. (2019). The Benefits of Linking Metadata for Internal and External Users of an Audiovisual Archive. En *MTSR 2018: Metadata and Semantic Research* (pp. 212–223). doi: 10.1007/978-3-030-14401-2_20
- Iglésias Franch, D. (2008). *La fotografía digital en los archivos*. Gijón: Trea.

- Lally, A.M. & Dunford, C. E. (2007). Using Wikipedia to Extend Digital Collections. *D-Lib Magazine*, 13(5/6). Recuperado de: <http://www.dlib.org/dlib/may07/lally/05lally.html>
- Navarrete, T. y Borowiecki, K.J. (2016). Changes in cultural consumption: ethnographic collections in Wikipedia. *Cultural Trends*, 25(4), pp. 233-248. doi: 10.1080/09548963.2016.1241342
- Phillips, L. B. (2013). The Temple and the Bazaar: Wikipedia as a Platform for Open Authority in Museums. *Curator: The Museum Journal*, 56(2): 219-235. doi:10.1111/cura.12021
- Robledano Arillo, J. & Canela Garayoa, M. (2007). Estándares para la descripción de fotografía. *Revista d'arxius*, 6, pp. 149-188
- Sánchez Vigil, J.M. (2006). *El documento fotográfico. Historia, usos, aplicaciones*. Gijón, Trea.
- Sánchez Vigil, J. M. & Salvador Benítez, A. (2014). *Documentación fotográfica*. Barcelona, UOC.
- Saorín, T. & Pastor-Sánchez, J.A. (2011). Bancos de imágenes para proyectos enciclopédicos: el caso de Wikimedia Commons. *El Profesional de la Información*, 20(4), 424-431. doi:10.3145/epi.2011.jul.09
- Szajewski, M. (2013). Using Wikipedia to Enhance the Visibility of Digitized Archival Assets. *D-Lib Magazine*, 19(3/4). Recuperado de: <http://www.dlib.org/dlib/march13/szajewski/03szajewski.html>
- Tramullas, J. (2015). Wikipedia como objeto de investigación. *Anuario ThinkEPI*, 9, 223-226. doi:10.3145/thinkepi.2015.50
- Tramullas, J. & Ojeda, R. (2018). Fondos documentales para el estudio de la historia de la fotografía en Wikimedia Commons. En: Hernández-Latas, J.A. (ed.) *II Jornadas sobre Investigación en Historia de la Fotografía* (pp. 407-418). Zaragoza: Institución Fernando el Católico.
- Tramullas, J., Sánchez-Casabón, A. I., & Garrido-Picazo, P. (2018). Wikipedia categories in research: towards a qualitative review of uses and applications. En: *Challenges and Opportunities for Knowledge Organization in the Digital Age. Proceedings of the Fifteenth International ISKO Conference*. (pp. 490-497). Porto: Ergon Verlag.
- Vrandečić, D., & Krötzsch, M. (2014). Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10), 78-85. doi: 10.1145/2629489