

Infraestructuras de datos e investigación para patrimonio cultural en Europa

Data and research infrastructures for cultural heritage in Europe

Gustavo Candela

Candela, G. (2026). Infraestructuras de datos e investigación para patrimonio cultural en Europa. *Anuario ThinkEPI*, 20, e20a11. <https://doi.org/10.3145/thinkepi.2026.e20a11>

Publicado en *IweTel* el 12 de mayo de 2026

Gustavo Candela

<https://www.directorioexit.info/ficha6986>

<https://orcid.org/0000-0001-6122-0777>

Universidad de Alicante

gcandela@ua.es



Resumen: Las infraestructuras de datos e investigación para patrimonio cultural a nivel europeo facilitan la ciencia abierta y la computación en la nube. Este trabajo pretende analizar las últimas tendencias y trabajos sobre el uso y publicación de datos en patrimonio cultural. Además, presenta las iniciativas que se están desarrollando actualmente en Europa. Se concluye con un análisis sobre las posibilidades y desafíos futuros.

Palabras clave: Infraestructuras europeas; Patrimonio cultural; Colecciones digitales.

Abstract: Data and research European infrastructures for cultural heritage facilitate open science and cloud computing. This work aims to analyze recent initiatives and studies on the use and publication of data in cultural heritage. It presents the initiatives currently being developed. It concludes with an analysis of the opportunities and challenges.

Keywords: European infrastructure; Cultural heritage; Digital collections.

1. Introducción

Las instituciones de patrimonio cultural han explorado durante los últimos años nuevas formas para poner a disposición del público sus colecciones en formato digital. Los avances en materia de tecnología han favorecido un nuevo contexto en el que los datos son un ingrediente esencial para los algoritmos basados en aprendizaje automático e inteligencia artificial. Las instituciones de patrimonio cultural llevan décadas catalogando sus registros y materiales de forma detallada, utilizando estándares internacionales y asegurando la calidad de la información (Lorenzini et al., 2021; Candela, 2023; Dişli et al., 2025a).

Entre las iniciativas más relevantes en materia de publicación de colecciones digitales que facilitan el acceso computacional se encuentra *Collections as data* (Padilla et al., 2023; Candela, 2024). Esta propuesta fue financiada por la Andrew W. Mellon Foundation y promueve el acceso a los contenidos en formato digital para aplicar técnicas como análisis de redes, visualización, enriquecimiento o visión por computador. Numerosas iniciativas han adoptado estos principios para la publicación de sus contenidos como, por ejemplo, el Laboratorio de Humanidades Digitales del Instituto de Investigaciones Bibliográficas y Crítica Textual en Argentina, la Biblioteca Nacional de España con su proyecto de datos abiertos y enlazados o la Biblioteca Nacional de Escocia y el Data Foundry (Candela, 2026a). Todos estos proyectos tienen en común la puesta a disposición del público de colecciones digitales que incluyen diferentes tipos de contenido y formatos, con el propósito de que puedan ser reutilizados por la comunidad.

En paralelo, se han creado varias comunidades internacionales enfocadas en el uso responsable de las colecciones digitales como Artificial Intelligence for Libraries, Archives & Museums (AI4LAM) (<https://ai4lam.org/>) y la International GLAM Labs Community (<https://www.glamlabs.io/>). Estas comunidades aspiran a dotar a las instituciones de patrimonio cultural de los conocimientos necesarios para poder explorar el uso de las nuevas tecnologías y mantener su relevancia en el nuevo contexto tecnológico.

Una de las principales herramientas que han utilizado las instituciones de patrimonio cultural para demostrar a los investigadores cómo reutilizar las colecciones digitales disponibles en forma de datos son ejemplos de código reproducible basados en Jupyter Notebooks (Candela et al., 2023). Este tipo de herramientas permite combinar texto, documentación, código, gráficas y componentes interactivos, siendo posible su ejecución en un simple navegador sin necesidad de instalar ningún software adicional. Una de las principales ventajas en materia de investigación consiste en que permiten reproducir los resultados obtenidos en un estudio. También se han convertido en elementos indispensables para talleres de formación y educación. Sin embargo, las plataformas gratuitas de ejecución tienen límites en cuanto a recursos como espacio en disco, uso de tarjetas gráficas y memoria. En otros casos, mayoritariamente basados en compañías privadas de tecnología, se proporcionan más recursos de este tipo, pero con el inconveniente que conlleva la compartición y soberanía de datos, así como también las diferencias en cuanto al interfaz y los estándares utilizados, que dificultan su reutilización en otras plataformas. Resulta fundamental a nivel europeo proporcionar un entorno que permita la ejecución de este tipo de herramientas, así como asegurar su sostenibilidad en el futuro.

Tradicionalmente la compartición de datos ha sido a través de plataformas de datos abiertos. Los avances en materia de inteligencia artificial durante los últimos años han generado situaciones en las que las instituciones perdían el control del uso de los datos. En este sentido, y para mitigar estos problemas, desde Europa se promueve la creación de infraestructuras federadas que facilitan nuevas formas de compartición de datos basadas en la soberanía y seguridad.

2. El contexto europeo

La iniciativa Europea es uno de los principales actores a nivel europeo de patrimonio cultural. Europea tradicionalmente ha agregado los proyectos de patrimonio cultural europeos en un único interfaz, a través de un mismo modelo de datos y el uso de APIs. Recientemente, la iniciativa Europea se ha convertido en la responsable de la implementación del espacio común europeo de datos de patrimonio cultural (Europeana Foundation, 2025). Un espacio de datos es un sistema federado que facilita la seguridad y soberanía de los datos. Cada participante puede tener un rol de tipo proveedor o consumidor, y se le provee de un conector para la publicación y consumo de datos. Este tipo de infraestructuras ha generado un nuevo paradigma para la compartición de datos

diferente a los sistemas tradicionales en los que las instituciones pierden el control del uso de los datos. También incluye un componente de gobernanza para asegurar el correcto funcionamiento de la infraestructura. En este sentido, la iniciativa Europea en colaboración con DARIAH (Digital Research Infrastructure for the Arts and Humanities) ha trabajado en el análisis para la adopción de los principios de *Collections as Data* en el contexto del espacio de datos de patrimonio cultural (Candela et al, 2025a). En el marco de Europeana Research and Tech Communities, se ha definido un grupo de trabajo para desarrollar normas estandarizadas para elaborar documentación detallada sobre trazabilidad, origen y procedencia de los datos, facilitando un vocabulario semántico con clases y propiedades para su descripción (Alkemade et al., 2025). También se han analizado las posibilidades de enriquecimiento con repositorios externos y colaborativos como el caso particular de Wikidata (Europeana Foundation, 2026; Wikimedia España, 2026). Estas actividades reflejan la necesidad de las instituciones por adoptar este tipo de propuestas y la integración con nuevas plataformas de compartición y enriquecimiento de datos, donde la Europea está jugando un papel fundamental.

La ECCCH en Europa se refiere a la European Collaborative Cloud for Cultural Heritage (en español: Nube Colaborativa Europea para el Patrimonio Cultural). Es una iniciativa de la Unión Europea que tiene como objetivo crear una infraestructura digital común para el sector del patrimonio cultural (ECCCH, 2025). Su objetivo principal es construir un ecosistema digital europeo común para el patrimonio cultural, facilitando el acceso, la interoperabilidad y la reutilización de datos culturales entre instituciones. Recientemente se han financiado diferentes proyectos basados en casos de uso para dotarla de contenido, servicios y aplicaciones.

Otra infraestructura adicional para la compartición de datos científicos es la European Open Science Cloud (EOSC) es una iniciativa de la Unión Europea para crear una infraestructura digital federada de ciencia abierta en Europa. La EOSC promueve la publicación de datos científicos y facilita un conjunto de servicios para computación en la nube entre los que se incluye el almacenamiento y ejecución de Jupyter Notebooks, la compartición de grandes cantidades de datos o la creación de máquinas virtuales (European Open Science Cloud - EU Node, 2025).

Este tipo de infraestructuras centradas en la computación en la nube facilitan la creación de flujos de trabajo reproducibles para la reutilización y análisis de datos. En este sentido, resulta fundamental favorecer la interoperabilidad entre estas infraestructuras para facilitar la comunicación entre ellas. Esto permite generar escenarios en los que los datos se recopilan de una infraestructura y son reutilizados en un servicio disponible en otra infraestructura de forma transparente para el usuario final. Nótese que actualmente se está trabajando en esta línea, que se contempla como unos de los principales objetivos y al mismo tiempo supone un desafío importante.

3. Ejemplos de aplicaciones

Esta sección pretende proporcionar una pequeña selección de ejemplos representativos centrados en la reutilización de colecciones digitales que facilitan su acceso computacional. Estos ejemplos se podrían implementar en este tipo de infraestructuras de datos e investigación fomentando la ciencia abierta y la reproducibilidad.

Gracias al uso de APIs es posible la integración de diferentes repositorios de datos. De esta forma se puede tener un visión unificada de los datos proporcionados por diferente repositorios. Por ejemplo, trabajos previos se han enfocado a la definición de casos de estudio que integran datos de diferentes repositorios como Wikidata y las bibliotecas nacionales de España y Francia (Dişli et al., 2025b). Esto permite análisis avanzados y a gran escala, evitando los silos de datos.

Otros ejemplos se han centrado en la ECCCH para definir una selección de casos de uso dedicados a diferentes temáticas como la inteligencia artificial, computación de alto rendimiento, ciencia abierta o uso de contenidos 3D. Por ejemplo, el uso de modelos de lenguaje de gran tamaño para la consulta de colecciones digitalizadas o la aplicación de técnicas de visión por computador en mapas para el enriquecimiento de los metadatos [Candela et al., 2025b].

Otros ejemplos se basan en la extracción de colecciones digitales incluyendo metadatos e imágenes sobre una temática en particular como la Guerra Civil Española o la Segunda Guerra Mundial [Candela, 2026b]. En estos casos se requiere además proveer de documentación adicional sobre el proceso de extracción y disseminación.

4. Conclusiones

Con el apoyo de la financiación de la Unión Europea se están desarrollando un conjunto de infraestructuras con un mismo objetivo centrado en la transparencia, seguridad y compartición de datos. Las instituciones de patrimonio cultural deben realizar un esfuerzo y adaptarse a las nuevas necesidades para mantener su relevancia en el contexto actual. Este tipo de infraestructuras facilita la reutilización de las colecciones y por tanto el incremento del impacto en la sociedad de esos contenidos que en muchos casos están escasamente utilizados.

Europa ha definido la forma en que se comparten los datos apostando por infraestructuras federadas. Sin embargo, todavía existen desafíos referentes a la interoperabilidad entre estas infraestructuras, y también otras relacionadas con la sostenibilidad y futuro.

5. Declaración sobre el uso de Inteligencia Artificial

El autor declara que no ha utilizado herramientas de inteligencia artificial generativa en ninguna fase de elaboración de la presente nota.

6. Agradecimientos

El autor agradece el apoyo y la colaboración de las instituciones y personas que han contribuido al desarrollo de este trabajo, muchas de ellas mencionadas directamente en el texto. Asimismo, se reconoce el apoyo proporcionado por los proyectos e infraestructuras de investigación relacionados con patrimonio cultural digital y ciencia abierta. Este trabajo ha sido financiado por el proyecto Sophia (TSI-100130-2024-10) financiado por el Ministerio Para la Transformación Digital y la Función Pública, en el Plan de Recuperación, Transformación y Resiliencia, financiado por la Unión Europea - NextGenerationEU.

7. Referencias

- Alkemade, H., Candela, G., Claeysens, S., Colavizza, G., Eren, S., Freire, N., Irollo, A., Isaac, A., Lehmann, J., Neudecker, C., Osti, G., van Strien, D., & Wevers, M. (2025). Datasheets for Digital Cultural Heritage Datasets (Version 2). *Zenodo*. <https://doi.org/10.5281/zenodo.15828222>
- Candela, G. (2023). An automatic data quality approach to assess semantic data from cultural heritage institutions. *Journal of the Association for Information Science and Technology*, 74(7), 866–878. <https://doi.org/10.1002/asi.24761>
- Candela, G. (2024). Collections as data: Acceso computacional a colecciones digitales. *Anuario ThinkEPI*, 18. <https://doi.org/10.3145/thinkepi.2024.e18a06>
- Candela, G. (2026a). Towards a semantic approach in GLAM Labs: The case of the Data Foundry at the National Library of Scotland. *Journal of Information Science*, 52(1), 3–21. <https://doi.org/10.1177/01655515231174386>
- Candela, G. (2026b). Data Publishing and Use in the Humanities: A Labs Approach. *Zenodo*. <https://doi.org/10.5281/zenodo.18959285>

- Candela, G., Chambers, S., & Sherratt, T. (2023). An approach to assess the quality of Jupyter projects published by GLAM institutions. *Journal of the Association for Information Science and Technology*, 74(13), 1550–1564. <https://doi.org/10.1002/asi.24835>
- Candela, G., Chambers, S., Irollo, A., Freire, N., Dritsou, V., Isaac, A., Benardou, A., Garnett, V., & Tasovac, T. (2025a). A workflow to publish Collections as Data: Looking back at Europeana.eu and forward to the common European data space for cultural heritage. *Transformations: A DARIAH Journal*, 1. <https://doi.org/10.46298/transformations.14774>
- Candela, G., Dobрева, M., Alkemade, H., Holownia, O., Mahey, M., Ames, S., Renaud, K., Vodopivec, I., Lee, B. C. G., Padilla, T., Claeysens, S., Huvila, I., & Knazook, B. (2025b). A use case lens on digital cultural heritage. *ACM Journal on Computing and Cultural Heritage*, Online AM. <https://doi.org/10.1145/3828548>
- Dişli, M., Candela, G., Gutiérrez, S., & Fontenelle, G. (2025a). Open data practices of art museums in Wikidata: A compliance assessment. *Journal of Open Humanities Data*, 11(1), 71. <https://doi.org/10.5334/johd.438>
- Dişli, M., Osti, G., Candela, G., & Zijdemann, R. (2025b). From linked open data to collections as data: A reproducible framework using federated queries. *Information Technology and Libraries*, 44(4). <https://doi.org/10.5860/ital.v44i4.17432>
- ECCCH - The Cultural Heritage Cloud. 2025. <https://www.echoes-eccch.eu/>
- Europeana Foundation. (2025). *Espacio común europeo de datos para el patrimonio cultural*. <https://www.dataspace-culturalheritage.eu/es>
- Europeana Foundation. (2026). *Wikidata y el ecosistema GLAM: ideas del evento Wikimedia España en el Museo del Prado*. <https://www.dataspace-culturalheritage.eu/es/news/wikidata-and-the-glam-ecosystem-insights-from-the-wikimedia-espana-event>
- European Open Science Cloud - EU Node. (2025). Interactive Notebooks. A shared space for coding and analysis. <https://open-science-cloud.ec.europa.eu/services/interactive-notebooks>
- Lorenzini, M., Rospocher, M., & Tonelli, S. (2021). On assessing metadata completeness in digital cultural heritage repositories. *Digital Scholarship in the Humanities*, 36 (Supplement_2), ii182–ii188. <https://doi.org/10.1093/lc/fqab036>
- Padilla, T., Scates-Kettler, H., Varner, S., & Shorish, Y. (2023). Vancouver Statement on Collections as Data. *Zenodo*. <https://doi.org/10.5281/zenodo.8342171>
- Wikimedia España. (2026). *Wikidata en el ecosistema GLAM: impulsado por Wikimedia España desde el Museo del Prado*. <https://wikimedia.es/patrimonio-conectado-wikidata-impulsa-el-ecosistema-glam-desde-el-museo-del-prado/>