

Diseño y desarrollo de un marco de trabajo para la publicación de colecciones digitales bajo principios de computación mínima

Deseño e desenvolvemento dun marco de traballo para a publicación de coleccións dixitais baixo principios de computación mínima

Design and development a framework for publishing digital collections under minimal computing

Alumno/a: **LORENA PITA CASTRILLÓN**

Director/a: **MIGUEL ÁNGEL RODRÍGUEZ LUACES**

Grado: **XESTIÓN DIXITAL DE INFORMACIÓN E DOCUMENTACIÓN**

Año académico: **2022-2023**

Convocatoria: **3ª (septiembre 2023)**

Resumen

La democratización de la tecnología ha revolucionado diversos ámbitos de la sociedad, incluyendo sectores tradicionales como museos, archivos y bibliotecas. Un factor clave en esta transformación ha sido la digitalización, que ha permitido un acceso sin precedentes a la consulta y estudio de documentos patrimoniales que, de otra forma, sería inimaginable. Sin embargo, la gestión y divulgación de estas colecciones digitales no está carente de importantes desafíos.

Este trabajo se centra en la pregunta fundamental de si es viable gestionar y compartir con éxito colecciones digitales patrimoniales sin requerir conocimientos técnicos avanzados o costosos recursos tecnológicos. Para abordar esta cuestión, se emplea una metodología que combina una revisión exhaustiva de literatura relacionada con la publicación en línea de colecciones digitales y la evaluación de las principales soluciones de *software* del mercado. Esta combinación de enfoques proporciona una comprensión integral de las soluciones disponibles y sienta las bases para el desarrollo de un marco de trabajo basado en el concepto de "computación mínima" que se oriente en simplificar la gestión, preservación y divulgación de colecciones digitales. A modo de prueba de concepto, se implementará la propuesta resultante en una colección de documentos históricos relacionados con la comarca de Ferrolterra.

La relevancia de esta investigación radica en la necesidad apremiante de hacer que el patrimonio cultural digitalizado sea accesible y fácilmente administrable para una audiencia diversa. En un contexto como el actual, marcado por la era digital y la conectividad, la capacidad de compartir y preservar nuestro legado cultural debe tornarse en una tarea esencial y alcanzable para los profesionales encargados de su gestión.

Palabras clave: patrimonio cultural, colección digital, GLAM, computación mínima, *software* libre, *CollectionBuilder*.

Resumo

A democratización da tecnoloxía revolucionou diversos ámbitos da sociedade, incluíndo sectores tradicionais como museos, arquivos e bibliotecas. Un factor chave nesta transformación foi a dixitalización, que permitiu un acceso sen precedentes á consulta e estudo de documentos patrimoniais que, doutro xeito, sería inimaxinable. Con todo, a xestión e divulgación destas coleccións dixitais non está carente de importantes desafíos.

Este traballo céntrase na pregunta fundamental de se é viable xestionar e compartir con éxito coleccións dixitais patrimoniais sen requirir coñecementos técnicos avanzados ou custosos recursos tecnolóxicos. Para abordar esta cuestión, emprégase unha metodoloxía que combina unha revisión exhaustiva de literatura relacionada coa publicación en liña de coleccións dixitais e a avaliación das principais solucións de *software* do mercado. Esta combinación de enfoques proporciona unha comprensión integral das solucións dispoñibles e senta as bases para o

desenvolvemento dun marco de traballo baseado no concepto de "computación mínima" que se oriente en simplificar a xestión, preservación e divulgación de coleccións dixitais. A modo de proba de concepto, implementarase a proposta resultante nunha colección de documentos históricos relacionados cos estaleiros de Ferrolterra.

A relevancia desta investigación radica na necesidade apremiante de facer que o patrimonio cultural dixitalizado sexa accesible e facilmente administrable para unha audiencia diversa. Nun contexto como o actual, marcado pola era dixital e a conectividade, a capacidade de compartir e preservar o noso legado cultural debe tornarse nunha tarefa esencial e alcanzable para os profesionais encargados da súa xestión.

Palabras chave: patrimonio cultural, colección dixital, GLAM, computación mínima, *software* libre, *CollectionBuilder*.

Abstract

The democratization of technology has revolutionized various sectors of society, including longstanding domains like museums, archives, and libraries. A pivotal driver in this transformation is digitization, enabling unprecedented access to heritage documents for consultation and study that would otherwise be unimaginable. However, the management and dissemination of these digital collections come with significant challenges.

This work centers on the fundamental question of whether it is possible to effectively manage and share heritage digital collections without necessitating advanced technical expertise or costly technological resources. To address this query, a methodology is employed that integrates a thorough review of literature pertaining to the online publication of digital collections and the assessment of leading software solutions available in the market. This amalgamation of approaches furnishes a comprehensive comprehension of the available solutions and lays the groundwork for developing a framework grounded in the concept of "minimal computing." This framework aims to streamline the management, preservation, and dissemination of digital collections. As a proof of concept, the resulting proposal will be put into practice with a collection of historical documents linked to the Ferrolterra shipyards.

The significance of this research is underscored by the urgent necessity to render digitized cultural heritage accessible and easily manageable for a diverse audience. In the contemporary context characterized by the digital era and connectivity, the ability to share and preserve our cultural legacy must become an essential and attainable task for the professionals responsible for its stewardship.

Keywords: cultural heritage, digital collection, GLAM, minimal computing, free software, *CollectionBuilder*.

Índice

1. Introducción	6
1.1. Pregunta de investigación	8
1.2. Objetivos	8
1.3. Justificación del trabajo	8
1.4. Metodología	8
2. Marco teórico	10
2.1. Instituciones de la memoria	10
2.2. Colección digital.....	12
2.3. Objeto digital.....	13
2.4. Formatos de archivos	14
2.5. Metadatos.....	17
2.5.1. Tipos de metadatos	19
2.5.2. Esquemas de metadatos	19
2.6. Modelos de datos	21
2.7. Formatos de serialización.....	23
2.8. Derechos autor.....	24
2.9. Políticas y criterios de selección	27
2.10. Computación mínima	29
3. Estado de la cuestión	31
3.1. Omeka S.....	32
3.2. CollectiveAccess	34
3.3. Greenstone	35
3.4. CollectionBuilder	36
3.5. Comparativa	39
4. Marco de trabajo propuesto	42
4.1. Especificaciones del producto final.....	42
4.2. Componentes	44
4.2.1. Plantilla del sitio web	45
4.2.2. Plantilla de metadatos	46
4.2.3. Constructor automático de la colección	47

4.3 Flujo de tareas.....	49
4.3.1. Generar una exposición	50
4.3.2. Editar una exposición	52
4.3.3. Abrir una exposición.....	52
4.3.4. Publicar una exposición	52
4.4. Prueba de concepto	54
7.4.1. Definición temática	54
4.4.2. Caracterización del <i>dataset</i>	54
4.4.3. Implementación.....	54
4.4.4. Producto resultante	55
5. Resultados y discusión	60
6. Conclusiones	62
7. Referencias bibliográficas	63

1. Introducción

En los últimos tiempos, se viene observado un notable interés por el patrimonio, su conservación y difusión, que abarca prácticamente todos los ámbitos. De este modo, el patrimonio, tradicionalmente apreciado por su singularidad, rareza y antigüedad, está evolucionando hacia un recurso de índole cultural y social. Su apropiada gestión y divulgación no solo permite fortalecer sus valores inherentes, sino que también fornece un entorno propicio para el desarrollo de una sociedad que posea un conocimiento más profundo de su propia historia y cultura e incluso para que ese patrimonio pueda ser el catalizador de nuevas investigaciones y descubrimientos¹.

No cabe duda que este cambio de perspectiva ha sido facilitado en gran medida por la consolidación y mejora de las Tecnologías de la Información y la Comunicación (TIC's), y más concretamente por la digitalización, un proceso técnico que ha ganado relevancia en las últimas décadas en el ámbito de las instituciones de la memoria, y que ha jugado un papel fundamental a la hora de democratizar el acceso al patrimonio cultural y, al mismo tiempo, permitir la conservación de los originales que contienen los principales hitos de la humanidad². No en vano, los esfuerzos de digitalización por parte de instituciones tanto públicas como privadas, ha ido en aumento en los últimos años³.

Sin embargo, este avance no ha estado exento de altibajos. A partir de 2008, la crisis económica afectó a los recursos humanos y materiales disponibles, ralentizando el progreso en este campo. A ello hay que sumar el hecho de que la gestión del patrimonio digital plantea importantes desafíos. Por una parte, la preservación a largo plazo de los datos digitales requiere la implementación de estrategias de almacenamiento y respaldo seguras, así como la continua actualización de formatos y tecnologías para evitar la obsolescencia. También está la cuestión de los derechos de autor, que debe ser abordada de manera adecuada para garantizar el uso ético y legal de los materiales digitalizados.

Por otra parte, aunque se han logrado desarrollar herramientas de difusión sumamente amigables e intuitivas para el usuario final, la tarea de publicar una colección de objetos digitalizados supone para sus gestores tomar difíciles decisiones debido a la abrumadora diversidad de opciones de

¹ Martín Torres Márquez. "Propuesta de "UCOMuseo virtual", museo y metamuseo de la Universidad de Córdoba (España)". *Revista de Formación e Innovación Educativa Universitaria* 5, n.º 3 (2012): 133-152. https://refiedu.webs.uvigo.es/Refiedu/Vol5_3/REFIEDU_5_3_1.pdf

² Luis Ángel García Melero. "La biblioteca digital revisitada". *Boletín de la ANABAD* (2009). http://eprints.rclis.org/13628/1/LA_BIBLIOTECA_DIGITAL_REVISITADA.pdf

³ Los primeros proyectos de digitalización masiva comenzaron a tomar forma a partir del año 1992. Algunos de estos proyectos pioneros fueron: CLASS (universidad de Cornell), *Project Open Book* (universidad de Yale) y *Making of America* (de las universidades de Michigan y Cornell). En el ámbito español cabe mencionar la *Biblioteca Digital Hispánica* y la *Biblioteca Virtual Miguel de Cervantes*, entre otros.

software disponibles en la actualidad. Cabe señalar, además, que la publicación exitosa de una colección digitalizada depende directamente de la capacidad de los sistemas elegidos para manejar los datos y garantizar su accesibilidad en línea, lo cual suele implicar costos asociados a su adquisición, actualización y mantenimiento. En este sentido, la interoperabilidad también juega un papel fundamental. La necesidad de que los diferentes componentes del software interactúen sin problemas y que la colección sea accesible en diversos dispositivos y plataformas (ya que la falta de una integración efectiva puede provocar una deficiente experiencia de usuario y limitar considerablemente el alcance de su difusión), a menudo se transforma en un reto complicado de resolver.

Pero eso no es todo, el proceso de publicación de una colección no solo implica la selección del hardware y el software más adecuados (en relación a las necesidades y recursos económicos disponibles), sino también que el personal que lo vaya a ejecutar cuente con una comprensión sólida de los conceptos informáticos y las técnicas de gestión de datos implicados en el proceso (manipulación de imágenes, preservación digital, diseño de interfaces de usuario, etc.), ya que la falta de estas habilidades puede obstaculizarlo enormemente, y la inversión de tiempo y esfuerzo necesarios para adquirirlas o desarrollarlas puede resultar abrumadora, especialmente para quienes solo están familiarizados con los aspectos curatoriales o históricos de la colección.

En definitiva, la publicación de una compilación que ha sido transformada a formato digital se encuentra ante obstáculos de gran relevancia debido a la amplia variedad de alternativas informáticas disponibles, a la imperante necesidad de contar con competencias en el ámbito de la informática, así como a la infraestructura técnica necesaria para llevar a cabo dicho proceso. La superación de estos retos demanda de una aproximación de carácter estratégico y de profesionales versados tanto en el ámbito de la tecnología como de la gestión documental. Todo ello con el fin último de brindar una experiencia de fácil acceso y enriquecedora para el usuario final. En consecuencia, resulta fundamental abordar de manera minuciosa y concertada cada uno de los aspectos implicados en el proceso de publicación de una colección digital, asegurando que la selección del software se ajuste a las necesidades específicas, y al mismo tiempo, garantizando que se cuente con el nivel adecuado de competencia informática a fin de poder ejecutar cada etapa de manera eficiente y efectiva. Asimismo, no se debe pasar por alto el aspecto infraestructural, ya que establecer un entorno propicio para el alojamiento y la distribución de la colección digitalizada demanda inversiones en términos de hardware y recursos tecnológicos.

En vista de estos desafíos, se torna claro que el enfoque requerido debe ser holístico y multidisciplinario, aunando conocimientos tanto del campo de las TIC's como de la materia curada. La confluencia de estas perspectivas se erige como la clave para sortear los obstáculos y culminar con éxito la creación de una experiencia en línea que resulte enriquecedora desde todos los ámbitos y que, al mismo tiempo, se encuentre al alcance de todos los usuarios interesados.

1.1. Pregunta de investigación

En el ámbito del patrimonio cultural, lo habitual es que las instituciones se vean obligadas a trabajar con *software* estándar, ya que el desarrollo de un sistema de gestión de colecciones personalizado simplemente no es económicamente viable. Tanto el uso de un modelo genérico estandarizado como el de un modelo específico altamente personalizado tienen un coste. En lo que respecta al *software* estándar, existen multitud de soluciones para publicar digitalmente colecciones patrimoniales, pero la mayoría requiere que el usuario cuente con conocimientos técnicos avanzados, o bien resultan muy costosas en términos de hardware y software.

En este contexto surge la pregunta de investigación: ¿Es posible gestionar y publicar con éxito una colección digital cuando se cuenta con restricciones significativas de conocimientos informáticos, así como de hardware y software?

1.2. Objetivos

El objetivo de esta investigación es diseñar y desarrollar un marco de trabajo, basado en principios de computación mínima, que permita convertir una colección de documentos patrimoniales digitalizados en un producto web altamente personalizable, a fin de facilitar y simplificar su gestión, preservación y publicación.

Asimismo, como prueba de concepto, se implementará dicho marco sobre una pequeña colección de documentos históricos relacionados con la comarca de Ferrolterra, con el fin último de demostrar la eficacia y utilidad del marco de trabajo propuesto.

1.3. Justificación del trabajo

Esta investigación se fundamenta en la necesidad de dar a conocer colecciones digitales que no están siendo divulgadas y puestas a disposición de la ciudadanía debido a que sus gestores o bien carecen de conocimientos informáticos avanzados, o bien sufren restricciones significativas de medios técnicos (hardware, software, etc.).

En este contexto, el desarrollo de un marco de trabajo adecuado tiene el potencial de democratizar el acceso y la gestión de patrimonio cultural digitalizado, lo que sin duda puede beneficiar a un amplio espectro de instituciones y comunidades interesadas en compartir su legado cultural de manera sencilla y efectiva.

Se trata, además, de una tarea que, a priori, se puede completar a un nivel satisfactorio dentro de los tiempos disponibles para la realización de este TFG y que puede seguir enriqueciéndose en el futuro prácticamente de manera ilimitada.

1.4. Metodología

El enfoque inicial del trabajo se centró en una revisión exhaustiva de la literatura existente en materia de publicación de colecciones digitales a fin de proporcionar una base teórica que permitiese comprender los conceptos clave relacionados.

Seguidamente, se llevó a cabo la consulta de trabajos sobre soluciones de *software* relacionadas con la publicación de colecciones digitales en línea que cumpliesen una serie de criterios establecidos previamente, tales como estar orientadas a la creación de repositorios o bibliotecas digitales, estar basadas en *software* libre y ser amparadas por una entidad de prestigio. Para ello, se optó por un enfoque cualitativo, basado en el estudio detallado de las características y funcionalidades ofrecidas por cada solución.

Además, para obtener una visión práctica y completa, se procedió a la instalación, uso y evaluación de cada una de las opciones seleccionadas. Esto implicó una interacción directa, permitiendo una exploración exhaustiva de sus interfaces, así como la realización de numerosas pruebas. Con ello, se logró obtener información muy valiosa acerca de la usabilidad de cada una de las soluciones analizadas, en un entorno operativo y funcional.

En resumen, la metodología utilizada en esta investigación combinó la revisión bibliográfica, el análisis de la documentación oficial y la experiencia práctica con las aplicaciones analizadas. Este enfoque integral permitió obtener una visión completa de las principales soluciones para la publicación de colecciones digitales en la web y sentar las bases para el desarrollo de un marco de trabajo que diese cumplimiento a los objetivos establecidos previamente.

Adicionalmente, se llevó a cabo una prueba de concepto del marco desarrollado, basada en un estudio de caso exploratorio, descriptivo y aplicado. Esto conllevó su implementación en un contexto real, a fin de poder verificar su eficacia y utilidad para la publicación de colecciones digitales en la web.

2. Marco teórico

A continuación, se tratará de establecer un marco teórico en el que se aborden de manera detallada y estructurada los conceptos esenciales que constituyen los cimientos de esta investigación. Este marco teórico servirá como base para el análisis subsiguiente del estado de la cuestión en materia de soluciones orientadas a la publicación de colecciones digitales, y proporcionará un entendimiento enriquecedor de los pilares conceptuales que las sustentan. Para ello, se abordarán todos aquellos elementos teóricos que se consideran fundamentales en el ámbito de la publicación de obras digitales patrimoniales, permitiendo así establecer un contexto que posteriormente guiará el desarrollo del marco de trabajo propuesto.

2.1. Instituciones de la memoria

Desde finales del siglo XX se comenzó a materializar una conciencia de colaboración por parte de bibliotecas, archivos y museos, que se ha puesto de relieve con la aparición de denominaciones que tratan de designar sus esfuerzos y proyectos conjuntos con el fin de preservar la herencia cultural de la humanidad. De este modo, apelativos como “instituciones de la memoria”, LAM (acrónimo de *Libraries, Archives and Museums*) o GLAM (*Galleries, Libraries, Archives and Museums*), buscan enfatizar aquello que une a estas unidades de información, en vez de lo que las diferencian. En español estos acrónimos han sido traducidos por BAM (Bibliotecas, Archivos y Museos) y BAMG (Bibliotecas, Archivos, Museos y Galerías), respectivamente⁴.

El primero de los términos, el de “instituciones de la memoria”, fue acuñado por Dempsey⁵ en un documento preparado para la Comisión Europea. Por su parte, el término LAM fue recogido por primera vez en el año 2008 en un informe de *OCLC Research*⁶, aunque en él se admite que la idea venía gestándose desde hacía tiempo.

No cabe duda de que los bibliotecarios, archiveros y museólogos son colectivos profesionales de tradiciones, finalidades y metodologías de trabajo muy diferentes, que desde su origen tendieron a crear compartimientos estancos, dando como resultado que apliquen tratamientos

⁴ Perla Olivia Rodríguez Reséndiz y María Teresa Fernández Bajón. “Conectando los saberes de bibliotecas, archivos y museos (BAM) en torno a la preservación de documentos analógicos y de origen digital”. (México: Universidad Nacional Autónoma De México, 2019).

https://ru.iibi.unam.mx/jspui/handle/IIBI_UNAM/L212

⁵ Dempsey, Lorcan. “Scientific, Industrial, and Cultural Heritage: A Shared Approach”. *Ariadne*, n.º 22 (1999). <http://www.ariadne.ac.uk/issue/22/dempsey/>

⁶ Diane M. Zorich, Giinter Waibel y Ricky Erway. *Beyond the Silos of the LAMs: Collaboration Among Libraries, Archives and Museums* (Ohio: OCLC Research, 2008).

<https://www.oclc.org/content/dam/research/publications/library/2008/2008-05.pdf>

documentales diferentes a materiales muy similares⁷. Así, por ejemplo, una fotografía puede acabar en un archivo, en una biblioteca o formar parte de la colección de un museo y, a pesar de tener el mismo soporte, recibirá un tratamiento diferente en cada una de estas instituciones.

En el caso de las bibliotecas, pioneras en la normalización del tratamiento documental, los fondos se suelen catalogar usando un conjunto de normativas encabezadas en la actualidad por RDA, MARC21 y el modelo conceptual FRBR⁸. En los años noventa se unieron archivos y museos a la creación de estándares para la descripción patrimonial, dando como resultado que los archivos utilicen como principales herramientas descriptivas de sus fondos EAD e ISAD G⁹, mientras que los museos utilizan diversos sistemas, siendo CDWA¹⁰ el más extendido¹¹.

Teniendo en cuenta que todas ellas preservan objetos digitales en diferentes formatos, lenguajes, tamaños y contenidos, es decir, sus colecciones digitales se caracterizan por ser poliformes y, además, no encontrarse fijas en un soporte permanente, sino que deben ser migradas de forma cíclica, se hace más evidente que los métodos, técnicas y tecnologías aplicadas a su tratamiento deben ser afines porque el eje de esta tarea es el objeto digital. Por ello, la gestión de contenidos digitales no es un ámbito exclusivo de ninguna de estas instituciones, sino común a todas ellas.

En la formación de esta conciencia de colectivo sin duda también contribuyó la población de Internet y la web en la década de los 90. Este hito abrió infinitas posibilidades a la difusión del

⁷ Teresa Jacinto y Joana Balsa de Pinho. "Instituciones de la Memoria. Reflexiones sobre sus particularidades". *Boletín de la ANABAD* 58, n.º4 (2008): 283-296. <https://www.anabad.org/wp-content/uploads/2011/03/2008.4.pdf>

⁸ El modelo conceptual FRBR establece la estructura de la catalogación bibliográfica, dividiendo los recursos en categorías como obra y manifestación. RDA (*Resource Description and Access*) utiliza los principios de FRBR para proporcionar pautas detalladas sobre cómo describir recursos. MARC21 es un formato de codificación ampliamente utilizado que puede incluir información de catalogación basada en RDA y FRBR, permitiendo la representación electrónica de registros bibliográficos coherentes y estructurados. En conjunto, FRBR, RDA y MARC21 trabajan en sinergia para mejorar la organización y accesibilidad de la información bibliográfica.

⁹ EAD (*Encoded Archival Description*) se centra en la descripción y estructuración de archivos electrónicos mientras que ISAD(G) (*International Standard Archival Description - General*) es un estándar más general que se aplica tanto a la descripción de archivos físicos como digitales.

¹⁰ CDWA (*Categories for the Description of Works of Art*) es un estándar utilizado en la catalogación y descripción de obras de arte. Proporciona un conjunto de categorías y términos estandarizados para identificar y describir sus elementos clave, tales como autor, título, fecha y material.

¹¹ Marina Salse, Núria Jornet y Javier Guallar. "El patrimonio universitario desde una perspectiva GLAM. Análisis de los sitios web de las universidades europeas". *Revista General de Información y Documentación* 31, n.º2 (2019): 521-543. <https://revistas.ucm.es/index.php/RGID/article/view/77215/4564456559308>

propio patrimonio, pero también evidenció la necesidad de poner en común estos esquemas originalmente separados y tender puentes que permitieran la interoperabilidad.

A raíz de los esfuerzos comunitarios llevados a cabo hasta el momento por todas estas instituciones se han podido crear grandes repositorios en línea del patrimonio nacional y mundial, entre los cuales cabe señalar la *World Digital Library*¹², *Europeana*¹³ o *Hispana*¹⁴.

2.2. Colección digital

El marco conceptual sobre el que se asientan las colecciones digitales fue definido por Khan y Wilensky en 1995¹⁵. En su trabajo trataron de establecer las entidades básicas que se encuentran en cualquier sistema en el que se almacena información en forma de objetos digitales, a la que se puede acceder libremente, como la gestionada por instituciones tales como bibliotecas, archivos y museos. Según estos autores todo sistema debe funcionar de la siguiente manera: una institución con material para ser divulgado construye con él un objeto digital; seguidamente el objeto se almacena en uno o más repositorios desde los cuales puede ser recuperado directamente o distribuido a los usuarios finales a través de proveedores de servicios.

Una colección digital se compone de elementos digitales cuidadosamente elegidos y organizados con el propósito de simplificar su acceso y utilización. Esta selección se establece previamente en una política de desarrollo de la colección, la cual debe ser acordada y documentada. Es esencial que cada colección esté detalladamente descrita, proporcionando al usuario información sobre sus características individuales, dimensiones, formato, restricciones de acceso, propiedad, así como cualquier detalle relevante para garantizar la autenticidad, integridad y comprensión de dicha colección.

Una colección de calidad se enriquece con metadatos para la descripción y administración de los elementos. Estos metadatos se aplican tanto a la colección en su totalidad como a cada elemento individual dentro de ella, o incluso pueden combinarse ambas opciones. Además de los elementos y los metadatos, una colección digital debe incluir al menos una interfaz que facilite el acceso, herramientas para buscar elementos, consultas basadas en campos clave, así como métodos para recuperar, visualizar y utilizar los objetos. Por lo general, las colecciones digitales son creadas por organizaciones o grupos de organizaciones en colaboración y a menudo como parte de un proyecto conjunto.

¹² <https://www.loc.gov/collections/world-digital-library/>

¹³ <https://www.europeana.eu/es>

¹⁴ <https://hispana.mcu.es/es/inicio/inicio.do>

¹⁵ Robert Kahn y Robert Wilensky. "A Framework for Distributed Digital Object Services" (1995).
<http://www.cnri.reston.va.us/k-w.html>

2.3. Objeto digital

Dentro del ámbito tecnológico complejo actual, se denominan "objetos digitales" a las unidades de información que poseen una naturaleza compleja y dinámica, y que se encuentran distribuidas en diversas ubicaciones de la red, cada una de ellas con una URI (Identificador Universal de Recursos) asociada. Estos objetos digitales tienen la capacidad de representar cualquier tipo de datos y pueden incluir la agregación de múltiples conjuntos de datos en un solo objeto digital. Están diseñados para representar relaciones entre ellos y requieren un sistema escalable y adaptable para su almacenamiento, gestión, acceso y reutilización, especialmente en entornos colaborativos, para mantenerse al día con las cambiantes tecnologías¹⁶.

Se pueden considerar dos tipos de objetos digitales: aquellos creados en base a otro objeto en formato analógico (libros impresos, manuscritos, piezas museísticas...) y los nacidos digitales (libros electrónicos, bases de datos, fotografías digitales, páginas web, ...).

Los objetos digitales deben digitalizarse o crearse en un formato de calidad que sea compatible tanto con las necesidades actuales como con las futuras y que permita la creación de copias de acceso para tales propósitos. Por lo tanto, deben ser accesibles en diferentes plataformas y seguir estándares o prácticas recomendadas que faciliten su uso e interoperabilidad. Además, deben perdurar con el tiempo, incluso cuando las tecnologías cambien, y deben ser desarrollados con características que fomenten la capacidad de interactuar con otros sistemas y la posibilidad de ser reutilizados.

En términos generales, los objetos digitales se caracterizan por su¹⁷:

- **Heterogeneidad:** esto significa que un objeto digital puede representar una amplia variedad de contenidos en diferentes formatos, incluyendo imágenes, libros electrónicos y los metadatos que los describen.
- **Complejidad:** pueden agrupar uno o más tipos de contenidos en un solo elemento digital. Estos contenidos pueden estar en diversos formatos y ubicaciones, y ser referenciados desde el objeto digital.
- **Dinamismo:** pueden permitir mecanismos de acceso que transforme dinámicamente los contenidos para un uso específico o contexto particular. Por ejemplo, mostrar una vista ampliada de una imagen o generar una tabla de contenidos en tiempo real.

¹⁶ Dayni Pérez-Hernández y Martha Dunia Delgado-Dapena. "Modelo de gestión de objetos digitales para la gestión de soluciones tecnológicas", *Ingeniería Industrial* 34, n.º1 (2013): 40-49. scielo.sld.cu/pdf/rii/v34n1/rii05113.pdf

¹⁷ Pérez-Hernández y Delgado-Dapena. "Modelo de gestión de objetos digitales..."

- **Interrelación:** pueden representar las relaciones entre diferentes objetos digitales, como "es parte de", "es miembro de", "contiene", así como las relaciones entre los contenidos dentro de un mismo objeto digital.

En resumen, el concepto de objetos digitales se refiere a las unidades individuales que componen las colecciones, es decir, son las piezas o elementos que conforman una colección específica. Es importante destacar que un mismo objeto puede formar parte de más de una colección, lo que significa que puede ser compartido y utilizado en diferentes contextos y conjuntos de elementos. Esta característica permite una mayor flexibilidad y la reutilización de los objetos en diversas colecciones, enriqueciendo así la variedad y el alcance de las mismas.

2.4. Formatos de archivos

Los formatos digitales actúan como el contenedor virtual que alberga la información. La finalidad de su gestión radica en garantizar que la información retenga todas sus características y pueda cumplir su propósito, ya sea proporcionar información, educación, entretenimiento o preservar la memoria¹⁸. Para lograrlo, es esencial emplear enfoques y metodologías que salvaguarden la calidad, accesibilidad y sustentabilidad de la información digital durante todo su ciclo de vida.

Algunos de los diversos formatos de salida que se pueden obtener a través de un tratamiento de digitalización son *Portable Document Format* (PDF), *Tagged Image File Format* (TIFF), *Joint Photographic Experts Group* (JPG), *Portable Network Graphic* (PNG), *Rich Text Format* (RTF) o *DOC/DOCX* (formatos nativos de *Microsoft Word*), si se emplean técnicas de OCR una vez manipulado computacionalmente el documento¹⁹.

A continuación, se comentan brevemente algunos de estos formatos:

- **PDF:** es un formato de almacenamiento de documentos compuesto (imagen vectorial, mapa de bits y texto), desarrollado por la empresa *Adobe*. Este formato permite diferentes grados de compresión, aunque siempre con pérdidas. Es un estándar muy extendido y mantiene una muy buena calidad, por lo que es un formato ideal para crear documentos secuenciales con varias páginas. El archivo PDF/A, se basa en un subconjunto del formato PDF, y es el estándar para la conservación a largo plazo de documentos electrónicos²⁰.

¹⁸ "File formats and standards", en *Digital Preservation Handbook*, 2.^a ed. (2015).

<https://www.dpconline.org/handbook/technical-solutions-and-tools/file-formats-and-standards>

¹⁹ Manuela Moro Cabero. "Identificación, caracterización y selección de formatos para la preservación del recurso digital". *Métodos de Información* 9, n.º16 (2018): 49-90. <http://dx.doi.org/10.5557/IIMEI9-N16-049090>

²⁰ UNE-ISO 19005-1:2008. Gestión de documentos. Formato de fichero de documento electrónico para la conservación a largo plazo. Parte 1: Uso del PDF 1.4 (PDF/A-1).

- **TIFF:** Es un formato de archivo digital que acepta compresión sin pérdidas, por lo que se conserva la información original completa. Proporciona gran calidad para imágenes en color, escala de grises y blanco y negro. Es un estándar de uso gratuito y el más utilizado con fines de conservación. Su principal inconveniente, el tamaño del archivo que genera²¹.
- **JPEG:** Se trata de un formato de alta calidad para imágenes en color que ofrece un alto grado de compresión, aunque siempre con pérdidas. Es un formato estándar ampliamente utilizado para visualización de pantalla y acceso rápido a la imagen²².

Con todo, los formatos van evolucionando conforme se van añadiendo nuevas capacidades. Los formatos más recientes, o incluso las versiones actualizadas de algunos ya existentes, pueden llevar a la obsolescencia de formatos previos a medida que las nuevas generaciones de software dejan gradualmente de admitir los más antiguos. La falta de compatibilidad de un software con formatos de archivo obsoletos puede resultar en la inutilidad de los datos. En este sentido, tanto los formatos de código abierto como los propietarios enfrentan el riesgo de obsolescencia: los proveedores ocasionalmente emplean la obsolescencia planificada para motivar a los clientes a actualizar a nuevos productos, mientras que las comunidades de software de código abierto pueden abandonar el soporte para formatos antiguos si así lo consideran oportuno. La obsolescencia también puede ser resultado de circunstancias ajenas al propio formato, como ocurre en las guerras comerciales²³.

No cabe duda de que la obsolescencia de los formatos de archivo representa un riesgo que debe ser asumida. No obstante, este problema podría no ser tan alarmante como lo percibía la comunidad de preservación digital hace aproximadamente una década. Muchos formatos de archivo establecidos siguen siendo relevantes, son compatibles y aún pueden utilizarse. Es muy probable que la mayoría de los formatos de archivo con los que se interactúa hoy en día sean ampliamente reconocidos y cuenten con buena compatibilidad en un futuro.

Cabe señalar que no todos los formatos digitales son idóneos ni están diseñados para su archivo o preservación. Por lo tanto, cualquier estrategia de conservación debe tener en cuenta los requisitos del contenido de la colección y tomar decisiones sobre el formato de archivo más

²¹ ISO 12234-2:2001. Electronic still-picture imaging - Removable memory - Part 2: TIFF/EP image data format.

²² ISO/IEC 10918-1:1994. Information technology - Digital compression and coding of continuous-tone still images: Requirements and guidelines.

²³ En los años 80 el formato de vídeo VHS se impuso a *Betamax* principalmente debido a factores comerciales y de mercado en lugar de a calidad técnica. VHS tenía una capacidad de grabación más larga, lo que lo hizo más atractivo para los consumidores. Además, la decisión de *Sony* de mantener el control sobre las licencias de *Betamax* limitó su adopción, mientras que VHS permitió una amplia variedad de fabricantes, lo que impulsó su popularidad.

adecuado para preservar esas cualidades²⁴. Correlacionar el contenido con una elección apropiada de formato para la conservación o el acceso resulta fundamental.

Existen diversos factores que deben tenerse en cuenta al seleccionar un formato de archivo adecuado. Estos factores aseguran que la elección sea compatible con los objetivos de conservación, acceso y uso a largo plazo. De acuerdo con las “Directrices para la preservación del patrimonio digital” publicadas en 2006 por la UNESCO²⁵, algunos de estos factores clave son:

- **Características del contenido:** la naturaleza del contenido ya sea texto, imagen, audio, video o un modelo en 3d, influye en el formato óptimo. Cada tipo de contenido puede requerir formatos específicos para mantener su calidad y funcionalidad.
- **Objetivos de uso:** es esencial establecer para qué se empleará el archivo. Ya sea para preservación a largo plazo, facilitar la edición o permitir su distribución para su visualización, los objetivos específicos influyen en la elección del formato adecuado.
- **Compatibilidad:** es vital seleccionar un formato compatible con una amplia gama de *software* y plataformas para garantizar que el contenido sea accesible en diferentes contextos y a lo largo del tiempo.
- **Calidad y compresión:** algunos formatos ofrecen mejores opciones de compresión que otros, lo que puede influir en la calidad visual o auditiva. La elección de un formato debe equilibrar la calidad con el tamaño del archivo. Los formatos con pérdida son aquellos en los que los datos se comprimen o se desechan como parte de la codificación. Así, por ejemplo, el formato MP3 es ampliamente utilizado para la distribución comercial de archivos de música a través de la web, ya que el proceso de codificación con pérdida da como resultado un tamaño de archivo más pequeño.
- **Metadatos:** los formatos que admiten metadatos incrustados pueden ser valiosos para describir y organizar el contenido, facilitando su búsqueda y gestión.
- **Longevidad y soporte:** optar por formatos establecidos y respaldados por la comunidad tecnológica es importante para evitar la obsolescencia y garantizar el soporte continuo.
- **Interoperabilidad:** si se espera que el archivo interactúe con otros sistemas o herramientas, es esencial elegir un formato compatible con las soluciones que se utilizarán.
- **Seguridad y privacidad:** si el contenido es sensible o requiere ciertas medidas de seguridad, el formato debe ser capaz de incorporar medidas de protección adecuadas.

²⁴ Manuela Moro Cabero. “Identificación, caracterización y selección de formatos...”

²⁵ UNESCO. “Directrices para la preservación del patrimonio digital” (2006).
https://unesdoc.unesco.org/ark:/48223/pf0000130071_spa

- **Estándares y buenas prácticas:** consultar estándares y recomendaciones de la industria puede proporcionar pautas valiosas para elegir un formato que cumpla con requisitos técnicos y de conservación.
- **Costos y recursos:** algunos formatos pueden requerir herramientas y recursos específicos para su creación y gestión. Evaluar los costes asociados es fundamental.
- **Futuro previsible:** considerar la evolución tecnológica y las tendencias de la industria puede ayudar a anticipar cambios en la viabilidad a largo plazo del formato.

En última instancia, la elección de un formato de archivo debe ser una decisión informada que equilibre las necesidades actuales y futuras del contenido con los factores técnicos, prácticos y de conservación²⁶.

2.5. Metadatos

Los metadatos se pueden definir como datos que ofrecen información sobre otros datos. En esencia, actúan como descriptores que agregan contexto y significado a la información subyacente. Para ilustrar este concepto, consideremos un libro en una biblioteca. El contenido del libro en sí mismo representa la información principal, pero elementos como la portada, el título, el autor y la fecha de publicación son ejemplos de metadatos que ofrecen información adicional sobre el libro, sin necesidad de leerlo en su totalidad.

En el mundo digital, los metadatos abarcan una variedad de aspectos, como la fecha de creación de un archivo, el autor, las etiquetas o palabras clave asociadas, el formato del archivo, su tamaño, ubicación, etc. Son como una etiqueta invisible que ayuda a clasificar, buscar y comprender la información en el contexto adecuado.

Indiscutiblemente, los metadatos desempeñan un papel fundamental en la extracción del valor real de la información. Sin su presencia, los datos se reducirían a meros caracteres o bits carentes de contexto y significado.

Existen múltiples razones que respaldan la esencialidad de los metadatos, algunas de las cuales se destacan a continuación:

- **Búsqueda efectiva:** los metadatos posibilitan una búsqueda precisa y eficaz de información al permitir el uso de etiquetas y palabras clave, lo que simplifica la localización

²⁶ Para saber más sobre los diferentes formatos en uso en la actualidad, se puede consultar el sitio de la *Library of Congress* sobre sustentabilidad de formatos digitales (<https://www.loc.gov/preservation/digital/formats/fdd/descriptions.shtml>), donde se recogen descripciones de formatos según tipos de registro (textual, imagen fija, imagen en movimiento, registros sonoros, etc.) e información de utilidad para cualquier persona o institución que está evaluando qué tipos de formato utilizar para fines de preservación.

de los datos necesarios sin la necesidad de revisar minuciosamente cada archivo individualmente.

- **Orden y categorización:** contribuyen a la organización y clasificación coherente de los datos, permitiendo la categorización de archivos según su tipo, temática, fecha u otros criterios relevantes.
- **Control de versiones:** son esenciales para rastrear las diferentes versiones y asegurar el uso de la más actualizada de un archivo.
- **Gestión de derechos y acceso:** también resultan fundamentales para el control de derechos de autor y el acceso a los datos, permitiendo conocer quién tiene permiso para acceder o modificar la información.
- **Auditoría y cumplimiento:** asimismo, son valiosos para fines de auditoría y cumplimiento, ya que posibilitan el seguimiento de quién ha accedido a los datos y cuándo lo ha hecho, información relevante en ciertos contextos legales o empresariales.
- **Contexto y comprensión:** proporcionan el contexto necesario para comprender plenamente los datos, ayudando a interpretar la intención detrás de los mismos y su relevancia en un contexto más amplio.

Los metadatos tienen aplicaciones en una amplia gama de sectores; en el ámbito de las instituciones de la memoria permiten a las bibliotecas, museos y archivos catalogar y organizar sus colecciones de manera eficiente, facilitando la búsqueda y recuperación de información. Sin embargo, a pesar de su importancia, trabajar con metadatos presenta algunos retos:

- **Consistencia:** mantener la consistencia en la aplicación de metadatos a lo largo del tiempo puede ser un desafío, especialmente en entornos colaborativos.
- **Actualización:** los metadatos deben actualizarse a medida que cambian los datos o las circunstancias. Olvidar actualizarlos puede llevar a información inexacta.
- **Privacidad:** algunos metadatos pueden contener información sensible. Es importante considerar cómo se manejará la privacidad en relación con los metadatos.
- **Complejidad técnica:** implementar sistemas que capturen y gestionen adecuadamente los metadatos puede ser técnicamente complejo.

2.5.1. Tipos de metadatos

Los metadatos pueden clasificarse en varias categorías según el tipo de información que describen²⁷:

Metadatos descriptivos

Ofrecen información que detalla el contenido y la finalidad de los datos, como títulos, descripciones, etiquetas, palabras clave y resúmenes. Estos metadatos son especialmente valiosos para llevar a cabo búsquedas eficientes y organizar la información de manera efectiva.

Metadatos administrativos

Se centran en la gestión y administración de los datos e incluyen datos sobre derechos de autor, permisos de acceso, fecha de creación, propietario del archivo, fecha de modificación y otros aspectos similares. Son fundamentales para el control y la seguridad de los datos.

Metadatos estructurales

Utilizados para describir la organización y las relaciones entre distintas partes de un conjunto de datos, especialmente en formatos como documentos XML, donde se especifica cómo se estructuran y conectan los elementos.

Metadatos técnicos

Suministran detalles técnicos acerca de los datos, como el formato del archivo, la resolución de imágenes, la configuración de audio, etc. Son esenciales para garantizar la compatibilidad y la correcta visualización de los datos.

Metadatos de uso

Realizan un seguimiento de cómo se ha empleado un conjunto de datos con el tiempo, incluyendo información sobre la frecuencia de acceso, la fecha de la última modificación, las personas que han consultado el conjunto, entre otros aspectos.

2.5.2. Esquemas de metadatos

Los esquemas de metadatos son sistemas estructurados de etiquetas y categorías que se utilizan para describir y organizar información de manera coherente. Funcionan como un conjunto de reglas y normas que estandarizan la forma en que se documenta la información, lo que facilita su

²⁷ Isabel Daudinot Founier. "Organización y recuperación de información en Internet: teoría de los metadatos". *ACIMED* 14, n.º 5 (2006). http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S1024-94352006000500006&lng=es&nrm=iso&tlng=es

búsqueda, organización y comprensión²⁸. Los esquemas de metadatos desempeñan un papel crucial en la gestión de datos, ya que permiten a las organizaciones y usuarios establecer un marco común para etiquetar y catalogar datos de manera consistente.

A continuación, se reseñan los principales esquemas de metadatos utilizados en la actualidad, cada uno de los cuales fue diseñado para satisfacer necesidades específicas de descripción y organización. Estos esquemas proporcionan directrices sobre qué metadatos deben incluirse, cómo deben estructurarse y qué estándares deben seguirse, asegurando que los datos se mantengan consistentes y sean interoperables, independientemente de la fuente o el contexto de uso. Esto es fundamental para garantizar una gestión efectiva de la información en diversos campos, desde bibliotecas y archivos, hasta la web y la ciencia de datos:

- **Dublin Core (DC)**: es un estándar de metadatos ampliamente utilizado para describir recursos digitales, como documentos en línea, imágenes y videos. Proporciona un conjunto básico de elementos descriptivos, como título, autor y fecha, que son útiles para la catalogación y recuperación de información en línea.
- **EAD (Encoded Archival Description)**: es un formato utilizado principalmente en archivos y bibliotecas para describir colecciones de documentos y archivos históricos. Permite la creación de registros estructurados que incluyen información sobre la organización, el contenido y la ubicación de los materiales.
- **MARC 21**: es un estándar de metadatos utilizado en bibliotecas para catalogar y gestionar colecciones bibliográficas. Ofrece un conjunto de campos y subcampos específicos para describir recursos bibliográficos, facilitando la organización y búsqueda de información en catálogos locales.
- **METS (Metadata Encoding and Transmission Standard)**: es un estándar utilizado en la preservación de objetos digitales y la gestión de contenido. Permite la estructuración de metadatos relacionados con objetos digitales y su contenido, facilitando la preservación a largo plazo y la presentación coherente de recursos.
- **PREMIS (Preservation Metadata: Implementation Strategies)**: diseñado para la preservación digital, se utiliza para registrar información sobre la administración, autenticidad y trazabilidad de objetos digitales a lo largo de su ciclo de vida, garantizando su integridad a largo plazo.
- **MODS (Metadata Object Description Schema)**: es un formato de metadatos utilizado para describir recursos bibliográficos y culturales en línea. Es más flexible que MARC y se

²⁸ Francisco Javier García Marco. "De la catalogación de patrimonio a los esquemas de metadatos", *Boletín de la ANABAD* 64, n.º4 (2014): 21-77. <https://www.anabad.org/wp-content/uploads/2014/11/2014.4.pdf>

utiliza en diversos contextos, incluyendo bibliotecas digitales y repositorios de objetos digitales.

- **TEI** (*Text Encoding Initiative*): es un estándar de metadatos utilizado para la codificación y marcado de documentos textuales, especialmente aquellos relacionados con la investigación académica y la edición textual. Permite una representación estructurada y enriquecida de textos para su análisis y presentación en línea.

En definitiva, los metadatos representan un componente esencial pero invisible en el contexto de la información digital. Desempeñan un papel crucial al proporcionar contexto, estructura y significado a los datos, lo que a su vez simplifica la tarea de buscar, gestionar y comprender la información de manera efectiva. Ya sea en la organización de una biblioteca o en la administración de una empresa, los metadatos encuentran aplicaciones en una amplia gama de sectores.

Conviene señalar que, a pesar de los posibles desafíos que puedan enfrentar, no se puede subestimar su importancia en la era digital. En última instancia, los metadatos representan la llave que desbloquea la riqueza subyacente en los datos, permitiendo un uso ágil y eficiente de la información. En un mundo cada vez más impulsado por los datos, los metadatos se convierten en una herramienta fundamental para aprovechar al máximo la información disponible.

2.6. Modelos de datos

Los modelos de datos son abstracciones fundamentales en el ámbito de la gestión de la información, ya que proporcionan una estructura y un contexto para organizar y relacionar los datos dentro de sistemas de almacenamiento o bases de datos. Cada modelo tiene sus propias características y ventajas, lo que permite adaptarse de manera óptima a diferentes tipos de datos y aplicaciones. Ya que sería imposible abordarlos todos en detalle, nos centraremos en aquellos más relevantes, permitiendo hacerse una idea de la diversidad y complejidad existente en este ámbito.

Modelo de datos relacional

Este modelo es uno de los más utilizados y reconocidos en el mundo de las bases de datos. Se basa en la idea de organizar los datos en tablas, donde cada tabla representa una entidad o concepto, y las filas de la tabla representan registros individuales relacionados con esa entidad. Las columnas, por su parte, definen los atributos o características de los registros. Para establecer relaciones entre diferentes entidades, se utilizan claves, que son valores únicos que identifican de manera inequívoca cada registro en una tabla.

Modelo de datos tabular

Los datos tabulares, son una implementación específica del modelo de datos relacional. Este formato es ampliamente utilizado en una variedad de aplicaciones, desde sistemas de gestión de bases de datos tradicionales hasta hojas de cálculo en *Excel* y archivos CSV.

Su estructura es especialmente útil en contextos donde se requiere una presentación ordenada y accesible de los datos, lo que facilita su interpretación y análisis. Además, debido a su simplicidad y uniformidad, los datos tabulares son altamente compatibles con una variedad de herramientas y plataformas de software, lo que hace que la migración y el intercambio de datos sean eficientes y efectivos.

Modelo de datos de documentos

A diferencia del modelo relacional, este enfoque utiliza documentos para almacenar información. Un documento es una unidad autónoma que contiene datos y metadatos, y puede ser estructurado o no estructurado. Los datos se almacenan en un formato que es fácilmente legible y comprensible, como JSON (*JavaScript Object Notation*). Este modelo es especialmente útil en bases de datos donde la flexibilidad en la estructura de los datos es esencial.

Modelo de datos de grafo

En este modelo, los datos se representan como nodos interconectados por relaciones. Los nodos representan entidades, y las relaciones representan las conexiones o interacciones entre estas entidades. Los modelos de datos de grafo son ideales para representar datos altamente interconectados, como las redes sociales, donde los usuarios (nodos) están conectados por amistades o interacciones (relaciones).

Modelo de datos jerárquico

Este modelo organiza los datos en una estructura de árbol, donde cada elemento tiene un padre y puede tener cero o más hijos. Es especialmente útil en aplicaciones que requieren una estructura de datos en cascada, como sistemas de gestión de contenido (CMS²⁹), donde los elementos jerárquicos se utilizan para organizar y representar la información de manera lógica.

En conclusión, la elección del modelo de datos adecuado depende de la naturaleza de los datos que se manejan y de los requisitos específicos de la aplicación. Cada modelo de datos tiene sus propias ventajas y limitaciones, y su selección debe basarse en una comprensión profunda de los datos y las necesidades del sistema en cuestión. La elección correcta del modelo de datos puede tener un impacto significativo en la eficiencia y la efectividad de un sistema de almacenamiento y gestión de datos.

²⁹ Los Sistemas de Gestión de Contenido (CMS, por sus siglas en inglés) facilitan la creación de sitios web, blogs y portales, incluso para aquellos sin conocimientos avanzados en programación. Ofrecen una variedad de características (personalización de diseños, gestión de usuarios, programación de publicaciones, etc.) lo que convierte la gestión de contenido en una tarea más ágil y sencilla.

2.7. Formatos de serialización

En el vasto mundo digital, la serialización de datos juega un papel esencial en la transferencia, almacenamiento y comunicación de información entre sistemas. Los datos, en su forma nativa, suelen ser estructuras complejas que necesitan ser convertidas en una representación que pueda ser transmitida y comprendida por otros sistemas y lenguajes. Es aquí donde entran en juego los formatos de serialización, facilitando la interacción eficiente y la interoperabilidad en un entorno tecnológico diverso y en constante evolución.

La serialización de datos es el proceso de convertir objetos o estructuras de datos en una secuencia de bytes que puede ser almacenada en un archivo, transmitida a través de la red o almacenada en una base de datos. Esta secuencia de bytes representa la información en una forma compacta y portátil, lo que permite su fácil transporte y manipulación.

Los formatos de serialización son las especificaciones prácticas que dictan cómo se representarán los datos en la secuencia de bytes. Aquí, exploraremos algunos de los formatos más utilizados:

- **JSON** (*JavaScript Object Notation*): Ampliamente conocido y utilizado, JSON es un formato ligero y legible por humanos que se basa en el modelo de objetos. Su estructura está compuesta por pares clave-valor y arrays, lo que lo hace adecuado para muchas aplicaciones web y APIs.
- **XML** (*eXtensible Markup Language*): Aunque más verboso que JSON, XML es altamente estructurado y permite describir datos jerárquicos y documentos complejos. Es comúnmente utilizado para intercambio de datos y configuración.
- **RDF** (*Resource Description Framework*): utilizado en la web semántica y en la representación de datos enlazados, describe recursos y sus relaciones mediante tripletas que incluyen sujetos, predicados y objetos. Es fundamental para la interoperabilidad de datos en la web y la representación semántica.
- **CSV** (*Comma-Separated Values*): Si bien es mucho más simple que los formatos anteriores, CSV es útil para representar tablas de datos en texto plano, donde cada línea del archivo representa una fila y los valores se separan por comas.

La elección del formato de serialización de datos depende de una serie de factores fundamentales que influyen en cómo los datos son representados y gestionados. Estos factores incluyen la eficiencia del formato, su legibilidad, la interoperabilidad, la complejidad de los datos, el tamaño del archivo serializado, el entorno en el que se utiliza y la disponibilidad de soporte y mantenimiento³⁰.

³⁰ “Modelos de datos y sus formatos de serialización”, *Docs: Museos Abiertos*.

<https://docs.museosabiertos.org/en/datos-vinculados/modelado/1-introduccion/1-5-modelos-de-datos-y-sus-formatos-de-serializacion>

En este sentido, la eficiencia del formato se refiere a la relación entre el tamaño del archivo serializado y la velocidad de serialización y deserialización. Algunos formatos están optimizados para minimizar el tamaño de los archivos, lo que resulta en una transferencia y almacenamiento más eficientes. Otros formatos priorizan la velocidad de codificación y decodificación de datos.

La legibilidad también es un aspecto importante, especialmente cuando se requiere que los datos sean comprensibles por los seres humanos. En aplicaciones donde la claridad y la legibilidad son esenciales, los formatos que facilitan la interpretación de datos por parte de las personas pueden ser preferibles.

Por su parte la interoperabilidad se relaciona con la capacidad de los datos para ser comprendidos y procesados por sistemas escritos en diferentes lenguajes de programación o tecnologías. En entornos donde se necesite compartir datos entre sistemas heterogéneos, la elección de un formato compatible con múltiples plataformas es crucial.

La complejidad de los datos también es un factor a considerar. Algunos formatos son ideales para datos simples y estructuras de datos básicas, mientras que otros son más adecuados para representar datos altamente estructurados y complejos, como documentos XML o datos enlazados RDF. El tamaño del archivo serializado puede ser un punto crítico en aplicaciones donde se busque minimizar el uso de almacenamiento o acelerar la velocidad de transferencia de datos, lo que lleva a la selección de formatos más compactos.

Además, el entorno y la industria en la que se utiliza un formato pueden influir en la elección. Algunos sectores pueden tener preferencias o estándares específicos para la representación de datos, lo que puede influir en la decisión.

Por último, el soporte y mantenimiento de un formato son aspectos esenciales para garantizar su fiabilidad a lo largo del tiempo. La existencia de una comunidad activa que respalde y evolucione el formato puede ser determinante en la elección final.

Con la continua evolución de la tecnología y la proliferación de nuevos casos de uso, es probable que aparezcan más formatos de serialización especializados para abordar necesidades específicas.

En resumen, los formatos de serialización de datos son las herramientas fundamentales que permiten a los sistemas de software comunicarse y compartir información en un mundo altamente conectado. La elección del formato adecuado depende de una variedad de factores que deben sopesarse cuidadosamente para lograr un equilibrio entre eficiencia, legibilidad y compatibilidad.

2.8. Derechos autor

Las nuevas formas de acceso al patrimonio, fruto de la incorporación de la tecnología digital al ámbito de las instituciones de la memoria, representan un reto para la legislación de derechos de autor en la era actual.

Antes de la era digital, la reproducción de obras presentaba pocos problemas, ya que los formatos eran principalmente físicos, como libros, manuscritos, revistas, fotografías, etc. Sin embargo, con la digitalización de obras y la creación de otras nuevas en formato digital, se ha vuelto esencial repensar los límites de la reproducción y distribución de contenido protegido por derechos de autor.

La Unión Europea ha estado atenta a estos cambios y ha promulgado normativas para abordar el acceso al patrimonio cultural y a obras protegidas en el entorno digital. La “Directiva 2001/29/CE del Parlamento Europeo y del Consejo, de 22 de mayo de 2001, relativa a la armonización de ciertos aspectos de los derechos de autor y derechos afines en la sociedad de la información” (DDASI)³¹ fue un punto de partida importante. Sin embargo, presentaba limitaciones significativas que obstaculizaron la digitalización del patrimonio cultural.

La DDASI estableció que bibliotecas, centros de enseñanza, museos y archivos podían beneficiarse de dos límites a los derechos de autor. El primero permitía la reproducción de obras protegidas sin fines comerciales. El segundo permitía poner a disposición del público, para fines de investigación o estudio personal, obras de sus colecciones, pero solo a través de terminales especializados en sus locales.

Aunque estas excepciones fueron un avance, tenían problemas fundamentales. En primer lugar, estas excepciones eran facultativas, lo que significa que no todos los Estados miembros las adoptaron de la misma manera, lo que llevó a una falta de armonización en las legislaciones nacionales. En segundo lugar, los términos eran demasiado restrictivos³². Por ejemplo, no permitían la realización de copias digitales de obras impresas para fines de conservación ni la digitalización masiva de obras protegidas por derechos de autor. Además, tampoco abordaban la accesibilidad transfronteriza en línea de estas obras.

La “Directiva 2019/790/ del Parlamento y del Consejo de 17 de abril de 2019 sobre derechos de autor y derechos afines en el mercado único digital” (DDAMUD)³³ marcó un hito crucial al abordar estas limitaciones y proponer soluciones que impulsarían un mercado único digital.

La DDAMUD establece excepciones obligatorias a los derechos de autor para permitir que las instituciones responsables del patrimonio cultural divulguen obras y otras prestaciones fuera del circuito comercial con fines no comerciales. Las condiciones para que esto sea posible incluyen que se mencione al autor o titular de derechos identificable y que las obras se pongan a disposición de forma gratuita. También se establece que una obra u otra prestación se considera "fuera del

³¹ DOUE-L-2001-81549, <https://www.boe.es/buscar/doc.php?id=DOUE-L-2001-81549>

³² María Serrano Fernández. “La digitalización del patrimonio cultural La Directiva sobre derechos de autor y derechos afines en el mercado único digital”. *Indret* 3 (2020). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3963849

³³ DOUE-L-2019-80826, <https://www.boe.es/buscar/doc.php?id=DOUE-L-2019-80826>

circuito comercial" cuando no está disponible al público a través de canales comerciales habituales después de un esfuerzo razonable para determinar su disponibilidad.

Sin embargo, surge una cuestión importante en relación con los metadatos asociados a estas obras. En principio, los metadatos que describen una obra, como fechas, nombres de autores y otros detalles, no están sujetos a derechos de autor, ya que no cumplen con el requisito de originalidad necesario para la protección de la propiedad intelectual. Sin embargo, los metadatos pueden contener información que, de manera individual, podría considerarse una obra protegida y, por lo tanto, requeriría el permiso del titular de los derechos. Por ejemplo, los metadatos que proporcionan un resumen del contenido de una obra, introducciones complementarias sobre el autor o contexto histórico, o una descripción detallada del formato de un volumen, podrían tener un grado mínimo de originalidad que cumpla con los requisitos de protección de derechos de autor.

Para ilustrar este punto, consideremos una biblioteca digital que ofrece obras digitalizadas. A menudo, esta biblioteca proporciona metadatos que acompañan a las obras, como resúmenes detallados o información sobre el autor. Estos metadatos pueden ser creados por especialistas y, en algunos casos, pueden cumplir con los requisitos de originalidad para la protección de derechos de autor. Esto plantea la cuestión de si el uso de estos metadatos requeriría el permiso de los titulares de derechos.

Este ejemplo resalta la complejidad de la protección de metadatos y cómo la DDAMUD y otras normativas deben abordar esta cuestión para garantizar un acceso adecuado al patrimonio cultural y las obras protegidas en el entorno digital.

Para abordar estos desafíos, es esencial encontrar un equilibrio entre el acceso al patrimonio cultural y la protección de los derechos de autor. Las excepciones y limitaciones establecidas en la DDAMUD son un paso importante hacia la armonización de las legislaciones en la UE y permiten un mayor acceso a obras y prestaciones fuera del circuito comercial. Sin embargo, la cuestión de los metadatos aún requiere una consideración cuidadosa.

Una posible solución podría ser distinguir entre metadatos que cumplen con los requisitos de originalidad y aquellos que son meramente descriptivos. Los metadatos descriptivos, como nombres de autores o fechas, podrían considerarse información funcional y no estar sujetos a derechos de autor. En cambio, los metadatos que alcanzan un nivel significativo de originalidad sí podrían estar sujetos a derechos de autor, aunque con limitaciones específicas para su uso en contextos culturales y educativos.

2.9. Políticas y criterios de selección

En 2007, la *National Information Standards Organization* (NISO) publicó un documento titulado "A framework of guidance for building good digital collections"³⁴, que estableció una serie de principios fundamentales que deben guiar la construcción de colecciones digitales de alta calidad. Los dos primeros principios son centrales para el presente estudio, ya que consideran las políticas de colección y la presentación que se hace de la colección a los usuarios, dos aspectos fundamentales que todas las colecciones digitales deberían tener en cuenta, pero que a menudo se pasan por alto.

Principio 1: Política de Desarrollo de la Colección Digital

El primer principio establecido por la NISO se refiere a la necesidad de contar con una política de desarrollo de la colección digital claramente definida antes de comenzar la construcción de la colección. Esta política debe ser el resultado de un acuerdo documentado y debe estar estrechamente alineada con la misión de la organización que la respalda.

La importancia de este principio radica en que proporciona una base sólida y coherente para la creación y gestión de la colección digital. Al enmarcar la política de desarrollo dentro de la misión de la organización, se asegura que la colección sirva a los objetivos y valores de la institución. Esto es esencial para garantizar que la colección digital tenga un propósito claro y beneficie tanto a los usuarios actuales como a los futuros.

La identificación de los usuarios a los que se dirige la colección es otro aspecto crucial de este principio. No se trata solo de considerar a los usuarios habituales, sino también de anticipar posibles usuarios no habituales o inesperados. Esto implica una visión a largo plazo y una comprensión profunda de las necesidades cambiantes de la comunidad a la que se sirve. Además, es importante destacar que la política de desarrollo debe estar relacionada con las políticas de colección generales si la institución también posee colecciones analógicas. Esto asegura una coherencia en la gestión de las colecciones en diferentes formatos.

Un elemento fundamental en este principio es la definición de criterios de selección para los materiales que se digitalizarán. Estos criterios deben estar bien establecidos y ser parte integral de la política de colección. Estos criterios de selección deben basarse en una planificación cuidadosa y deben reflejar los objetivos y necesidades de la comunidad de usuarios. En resumen, este principio establece las bases para una toma de decisiones informada y coherente en la construcción de colecciones digitales.

Principio 2: Descripción de las Colecciones

El segundo principio de la NISO se centra en la importancia de describir adecuadamente las colecciones digitales para que los usuarios puedan descubrir sus características clave. Esta

³⁴ NISO. *A Framework of Guidance for Building Good Digital Collections*. 3.^a ed. Baltimore: NISO, 2007.
<https://www.niso.org/sites/default/files/2017-08/framework3.pdf>

descripción debe incluir elementos esenciales como el alcance, el formato, las restricciones de acceso, la propiedad y cualquier información relevante para determinar la autenticidad, integridad e interpretación de la colección.

La descripción de la colección cumple dos funciones fundamentales. En primer lugar, facilita el acceso a la colección al proporcionar a los usuarios información que les permite comprender de qué se trata y si es relevante para sus necesidades. Esto es crucial en un entorno digital en el que la cantidad de información disponible puede resultar abrumadora. Una descripción completa y precisa permite a los usuarios encontrar y acceder a la información que están buscando de manera más eficiente. En segundo lugar, la descripción de la colección tiene como objetivo ayudar a los usuarios a comprender la colección en profundidad.

Entre la información que se puede ofrecer a los usuarios se encuentra la historia de la colección, su alcance o cobertura, los tipos de documentos que contiene y las condiciones de uso y reproducción. Esto no solo da una visión sobre lo que se puede esperar de la colección, sino que también ayuda a comprender su valor y relevancia en el contexto de la institución. Por ejemplo, si una biblioteca digital tiene una colección de manuscritos históricos, la descripción de la colección podría incluir detalles sobre la procedencia de estos manuscritos, su importancia en la investigación histórica y las políticas de acceso para proteger su preservación a largo plazo. Esta información no solo educa a los usuarios sobre la colección, sino que también puede despertar su interés y fomentar la participación en actividades de investigación y exploración.

Además de estos elementos, también se señala la utilidad de incluir información técnica relevante, como el software utilizado y los esquemas de metadatos. Esto puede ser especialmente importante para usuarios con conocimientos técnicos que deseen utilizar la colección de manera más avanzada.

En cuanto a los estándares y directrices para la elaboración de estas descripciones, el documento de la NISO señala que no hay un estándar dominante, pero menciona algunas propuestas, como el borrador de la norma NISO Z39.91³⁵ y las normas ISAD(G) y EAD. Estas normas proporcionan estructuras y formatos que pueden ser utilizados para crear descripciones coherentes y útiles de las colecciones digitales.

En resumen, los dos primeros principios establecidos por la NISO en su marco de orientación para la construcción de colecciones digitales son fundamentales para garantizar la calidad y la utilidad de estas colecciones. La creación de una política de desarrollo de la colección clara y bien documentada proporciona una base sólida para la toma de decisiones informadas, alineadas con

³⁵ NISO Z39.91 es un estándar desarrollado por el *National Information Standards Organization* (NISO) que se enfoca en la creación y mantenimiento de registros de metadatos bibliográficos para recursos electrónicos. Este estándar proporciona pautas y directrices para la descripción de recursos electrónicos, incluyendo libros electrónicos, revistas en línea y otros materiales digitales, con el objetivo de asegurar la coherencia y calidad en la catalogación.

la misión de la institución y las necesidades de los usuarios. Por su parte, la descripción detallada de las colecciones asegura que los usuarios puedan acceder a la información de manera eficiente y comprendan el valor y el contexto de la colección. Además, estos principios también desempeñan un papel importante en la comunicación con los usuarios, permitiendo que la institución comparta su visión y propósito en la creación de colecciones digitales.

No cabe duda de que la aplicación de estos principios es esencial para garantizar que las colecciones digitales sean valiosas, accesibles y perdurables a lo largo del tiempo. La adhesión a estos principios no solo beneficia a los usuarios actuales, sino que también contribuye a preservar y compartir el patrimonio cultural para las generaciones futuras.

2.10. Computación mínima

La "computación mínima" es un concepto que ha cobrado una creciente importancia en el campo de las humanidades digitales. Se trata de un enfoque que se centra en la implementación y el uso de la tecnología de la manera más esencial y básica posible para llevar a cabo tareas o funciones específicas³⁶. En lugar de recurrir a sistemas informáticos avanzados y complejos, la computación mínima busca aprovechar al máximo recursos limitados para lograr resultados prácticos y eficientes.

En este contexto, se da prioridad a la eficiencia y la simplicidad. Esto implica la eliminación de características innecesarias y el uso de recursos de manera óptima. Entre las prácticas comunes se incluye el uso de hardware de bajo costo, sistemas operativos ligeros y aplicaciones esenciales que se ajustan perfectamente a los requisitos específicos de la tarea en cuestión. Esta filosofía de "hacer más con menos" ha encontrado su aplicación en diversos contextos, desde dispositivos embebidos y sistemas de control industrial hasta sensores y actuadores en el "Internet de las Cosas" (IoT), así como en proyectos con presupuestos limitados.

La relevancia de la computación mínima radica en su capacidad para abordar desafíos específicos con una huella mínima. A pesar de las limitaciones en términos de potencia de cálculo, memoria y almacenamiento, la computación mínima se ha convertido en una opción atractiva en situaciones donde la eficiencia y la simplicidad son más importantes que la potencia y la versatilidad. Los desarrolladores que se adentran en el mundo de la computación mínima se enfrentan al desafío de optimizar sus programas y aplicaciones para que funcionen de manera eficaz en entornos con recursos limitados. Esto puede implicar el uso de algoritmos más eficientes, la reducción del consumo de memoria y la adaptación de las aplicaciones para que se ejecuten de manera fluida en plataformas con recursos limitados.

³⁶ Olivia Wikle. "What is static web and what's it doing in the digital humanities classroom?", *Dh Lib* (blog), 22 de junio de 2020. <https://dhandlib.org/2020/06/22/what-is-static-web-and-whats-it-doing-in-the-digital-humanities-classroom/>.

Un área en la que la computación mínima ha tenido un impacto significativo es en la publicación de colecciones de documentos digitalizados. Este enfoque representa un paradigma crucial para asegurar la accesibilidad universal a través de dispositivos con recursos limitados. Uno de los destacados enfoques en esta área es conocido como *Lib-Static*³⁷, que aboga por la simplicidad y la eficiencia en la presentación y distribución de contenido digitalizado.

El enfoque *Lib-Static* se basa en la creación de interfaces y plataformas de visualización altamente optimizadas. Los documentos se transforman en formatos ligeros y comprimidos, lo que reduce la carga de ancho de banda y el almacenamiento necesario. Esto, a su vez, permite un acceso fluido y rápido, incluso en dispositivos con conectividad limitada. La experiencia del usuario es fundamental en este enfoque, ya que se prioriza la simplicidad en la navegación y se minimiza la latencia en la respuesta, lo que facilita la interacción sin importar la capacidad del dispositivo utilizado.

Otro aspecto crucial de *Lib-Static* es la preservación del contenido a largo plazo. Al emplear formatos y técnicas duraderas, se garantiza que los documentos seguirán siendo accesibles y legibles en el futuro, lo que contribuye a preservar la integridad histórica y cultural de las colecciones. Este enfoque no solo trata de hacer que el contenido esté disponible en el presente, sino también de asegurarse de que pueda ser disfrutado y utilizado por las generaciones futuras.

En resumen, la computación mínima y el enfoque *Lib-Static* son dos conceptos que han revolucionado la forma en la que se aborda la tecnología informática en situaciones donde los recursos son limitados. Estos enfoques no solo han simplificado la experiencia del usuario y optimizado la distribución de contenido digitalizado, sino que también han garantizado la preservación a largo plazo de la información cultural e histórica. En un mundo en el que la tecnología sigue avanzando a pasos agigantados, la computación mínima representa un enfoque valioso y sostenible para abordar desafíos tecnológicos y llevar a cabo tareas esenciales de manera eficiente y efectiva.

³⁷ <https://lib-static.github.io/about/>

3. Estado de la cuestión

Una colección digital puede ser simplemente un listado de objetos digitales agrupados en una página web siguiendo algún orden simple, por ejemplo, el alfabético. Pero cuando la colección crece por encima de un determinado umbral, un simple listado deja de satisfacer las necesidades de los usuarios a la hora de encontrar obras de su interés, y asimismo se vuelve muy difícil de gestionar.

Es por ello que las colecciones digitales se gestionan mediante sistemas informáticos que permiten almacenar, ordenar, catalogar, poner a disposición del público y encontrar los objetos digitales. Mediante este sistema, los archivos digitales se cargan una única vez desde una interfaz, son adjudicados a colecciones, se les agrega la información, así como los metadatos correspondientes y se publican. De este modo, cada obra tendrá una única URL y quedará asociada a distintas categorías de información, como autor, año, género, colección, etc. Los repositorios por lo general incorporan uno o más sistemas estándar para asignar metadatos (como por ejemplo *Dublin Core*), por lo que no hace falta crear desde cero un sistema de catalogación propio.

En el ámbito de las instituciones de la memoria, las entidades normalmente se encuentran con la necesidad de emplear aplicaciones informáticas convencionales, dado que la creación de un sistema de administración de colecciones a medida generalmente resulta poco factible desde un punto de vista económico. El inconveniente de trabajar con software existente es que las instituciones a menudo se encuentran limitadas en cómo pueden describir sus objetos. Esto implica que el software de gestión de colecciones ya prescribe una determinada perspectiva, al utilizar un modelo preestablecido. Por lo tanto, no siempre es posible adaptarse a los requisitos específicos de una institución y sus colecciones, lo que puede llegar a generar frustración entre sus gestores.

Para construir un repositorio se puede optar a grandes rasgos entre dos opciones. La primera consiste en hacer uso de un sistema de gestión de contenidos web genérico (o CMS, por sus siglas en inglés), como *WordPress* o *Drupal*³⁸, que proveen las herramientas básicas de almacenamiento de archivos, gestión de contenidos, búsquedas, etc. La ventaja es que son

³⁸ WordPress y Drupal son dos de las plataformas de gestión de contenido más populares en el ámbito de la creación de sitios web. WordPress destaca por su facilidad de uso y versatilidad, lo que lo hace ideal para bloggers y pequeñas empresas que desean crear un sitio web de manera rápida y sencilla. Ofrece una amplia variedad de temas y complementos que permiten personalizar el diseño y las funcionalidades del sitio. Por su parte Drupal es una opción más potente y escalable que se adapta bien a sitios web más grandes y complejos. Aunque su curva de aprendizaje es más pronunciada en comparación con WordPress, brinda un mayor control y flexibilidad para desarrolladores y organizaciones que buscan una plataforma altamente personalizable. Es por ello que Drupal se utiliza comúnmente para crear sitios web institucionales, corporativos y comunidades en línea que requieren una estructura y seguridad robustas.

programas ampliamente usados, con millones de instalaciones en todo el mundo, comunidades que brindan soporte y actualizaciones permanentes. La desventaja es que no están diseñados específicamente para repositorios, y por lo tanto para algunas funciones avanzadas hay que agregar y configurar extensiones que no vienen por defecto, lo cual implica poseer unos ciertos conocimientos técnicos.

La segunda opción es utilizar un sistema de gestión específico para repositorios digitales. Estas herramientas ya vienen preparadas con todas las funciones para cargar, catalogar y visualizar colecciones digitales. Existen decenas de programas que podrían responder a esta definición, lo cual sin duda es una cantidad significativamente alta, complicando enormemente la elección.

En vista de lo anterior, se hace necesario restringir el número de soluciones a analizar en esta investigación; de este modo, para llevar a cabo la selección se han tenido en cuenta solo aquellas que cumpliesen las siguientes características:

- Estar orientada a la creación de repositorios y bibliotecas digitales.
- Ser y estar basadas en *software* libre o de código abierto.
- Con fecha de última actualización reciente.
- Gestionada o mantenida por una entidad de reconocido prestigio.

Sobre la base de estos criterios, se han seleccionado un total de cuatro aplicaciones: *Omeka S*, *CollectiveAcces*, *Greenstone* y *CollectionBuilder*.

3.1. Omeka S

Omeka S es una plataforma de código abierto diseñada para la creación, gestión y publicación de colecciones digitales y exposiciones en línea, desarrollada por la *Corporation for Digital Scholarship* del *Roy Rosenzweig Center* de la *George Mason University*, responsables también de otros proyectos como *Omeka Classic* o el sistema de gestión de referencias bibliográficas *Zotero*.

Se diferencia notablemente de otros sistemas por su enfoque basado en ontologías en lugar de utilizar vocabularios basados en XML. Esta elección de diseño tiene importantes implicaciones para la descripción de objetos en el repositorio y la representación de relaciones entre ellos y sus propiedades. En contraste con las especificaciones implementadas mediante XML, que suelen depender de un modelo jerárquico para representar estas relaciones, *Omeka S* se apoya en un modelo de grafos inspirado en el Framework de Descripción de Recursos (*Resource Description Framework* o RDF, por sus siglas en inglés). Este enfoque se alinea con los principios de los datos enlazados (*linked data*), que promueven la interconexión de información en la web de una manera semántica y significativa. Una de las ventajas clave de este enfoque radica en su capacidad para describir y conectar objetos en el repositorio de una manera más rica y flexible. En lugar de estar limitado por una jerarquía predefinida, *Omeka S* permite representar relaciones complejas y contextualizar la información de manera más precisa. Esto es esencial cuando se trabaja con colecciones digitales que contienen una variedad de tipos de objetos y metadatos. No obstante,

los datos se almacenan en una base de datos relacional *MySQL* incompatible con una interrogación directa mediante consultas *SPARQL* sobre los grafos RDF, lo cual supone una curiosa contradicción para una aplicación que apuesta por estas tecnologías³⁹.

Cabe señalar que también ofrece una serie de vocabularios precargados que son fundamentales para la descripción de objetos. Estos incluyen *Dublin Core*, *Dublin Core Type*, *Bibliographic Ontology* y *Friend of a Friend*. Estos vocabularios proporcionan un marco sólido para la organización de metadatos y la descripción de objetos, y son ampliamente reconocidos en la comunidad de bibliotecas y archivos. Además de estos vocabularios precargados, permite importar vocabularios adicionales en alguno de los formatos de serialización compatibles (*JSON-LD*, *N-Triples*, *Notation3*, *RDF/XML* o *Turtle*), con el fin de que las instituciones u organizaciones puedan satisfacer sus necesidades específicas utilizando vocabularios que sean relevantes para su contenido y contexto.

En lo que respecta a las funcionalidades, cuenta con una opción denominada "Sitios" que permite crear y gestionar diferentes instancias desde donde mostrar públicamente el contenido del repositorio, de tal modo que cada uno de estos entornos dispondrá de su propia URL y contará con sus propias páginas, sistemas de navegación, permisos de usuario y apariencia.

Aunque por defecto tan solo se incorporan un pequeño conjunto de características básicas, cuenta con un catálogo de módulos que pueden incorporarse al núcleo de la aplicación para ampliar sus funcionalidades básicas.

En cuanto a la interoperabilidad, *Omeka S* también dispone de un módulo específico para implementar el protocolo *OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting)*, si bien, actualmente solo es posible exponer metadatos y no recolectarlos desde terceras fuentes.

Si bien existen tanto proyectos de humanidades digitales como repositorios de bibliotecas y archivos basados en *Omeka S*, su uso destaca especialmente entre los primeros.

Dentro del panorama diverso y dinámico de proyectos que emplean *Omeka S* como plataforma de publicación, podemos destacar el proyecto *Cartesian cosmological illustrations*⁴⁰ de la universidad de Bucarest, *Ontologie du christianisme médiéval en images* del *Institut National d'Histoire de l'Art*⁴¹ o el *Museo Virtual* de la *Universitat de Barcelona*⁴², que ejemplifican de

³⁹ SPARQL y MySQL son tecnologías con propósitos diferentes y no son directamente compatibles. SPARQL es un lenguaje de consulta diseñado específicamente para bases de datos semánticas que utilizan el modelo RDF para representar datos en la web semántica, mientras que MySQL es un sistema de gestión de bases de datos relacionales que se utiliza principalmente para almacenar y consultar datos.

⁴⁰ <https://cartesian.unibuc.ro/s/cosmologicalillustrations/page/welcome>

⁴¹ <https://omci.inha.fr/s/ocmi/page/accueil>

⁴² <https://museuvirtual.ub.edu/>

manera elocuente la versatilidad y la amplitud de aplicaciones de esta solución.

3.2. CollectiveAccess

Es una plataforma de gestión de colecciones y activos digitales de código abierto diseñada para satisfacer las necesidades de museos, bibliotecas, archivos y otras instituciones culturales. Su origen, se remonta al año 2003, aunque su primera versión estable no se liberó hasta 2007, bajo el nombre de *OpenCollection*; fue en 2008 cuando adoptó su nombre actual. Desarrollado por la organización Whirl-i-Gig, cuenta con la colaboración de diferentes instituciones como *el Institute of Museum and Library Services*, *el National Endowment for the Humanities*, *el New York State Council for the Arts* o *el Kulturstiftung des Bundes*, entre otros.

CollectiveAccess se encuentra formado por dos componentes de software independientes: *Providence* y *Pawtucket*, encargados respectivamente de la parte de gestión y de difusión de los contenidos. *Providence* lo forman la base de datos, el entorno de trabajo mediante el cual gestionar los archivos digitales y una interfaz de usuario para la catalogación, búsqueda y gestión de las colecciones y los objetos digitales del repositorio. Cualquier instalación de *Collective Access* necesita, por tanto, disponer de una instancia de *Providence*. Por su parte, *Pawtucket* es el componente encargado de proporcionar un *frontend* o interfaz pública, si bien es posible utilizar cualquier otro que sea capaz de obtener información de la base de datos del sistema⁴³. La plantilla por defecto es muy simple, pero puede ser adaptada y personalizada modificando los correspondientes archivos.

Una de sus características más destacadas es su flexibilidad. La plataforma está diseñada para adaptarse a una amplia gama de tipos de colecciones, desde objetos físicos en museos hasta archivos digitales, bibliotecas y registros de investigación. Esto significa que las instituciones pueden utilizarlo para gestionar sus colecciones de manera integral, lo que facilita la creación de un repositorio unificado y la búsqueda de objetos a través de diferentes tipos de medios.

Otra de sus fortalezas reside en su capacidad para gestionar y describir metadatos de manera detallada y personalizable. Permite escoger entre un amplio abanico de esquemas de metadatos (*Dublin Core*, *DACS*, *PB Core*, *SPECTRUM*, *PREMIS*, etc.) o, incluso, crear uno completamente nuevo desde cero, así como definir los tipos de objetos digitales que formarán del repositorio. La plataforma también ofrece herramientas avanzadas para la importación y exportación de datos, lo que facilita la migración desde sistemas existentes y la compartición de datos con otras instituciones, lo cual resulta fundamental para la interoperabilidad y colaboración entre organizaciones culturales.

La personalización es otra característica sobresaliente de *CollectiveAccess*. Los usuarios pueden

⁴³ Un ejemplo es el *James Ensor Online Museum*

(<http://jamesensor.vlaamsekunstcollectie.be/en/collection>), en el que converge con *Drupal* o *el Religieus Erfgoed Online* (<http://www.religieuserfgoed.be/>), en el que se usa *Omeka* como *frontend*.

adaptar la plataforma para reflejar la estructura y los procesos de su institución, lo que permite una experiencia de gestión de colecciones altamente específica y orientada a las necesidades de cada usuario. También permite gestionar la información multilingüe y multicultural, lo cual resulta ideal para instituciones que trabajan con colecciones de diferentes idiomas y culturas.

CollectiveAccess también es reconocido por su enfoque en la colaboración y la comunidad. La plataforma cuenta con una comunidad activa de usuarios y desarrolladores que comparten recursos, ideas y soluciones. Esto facilita el soporte técnico y la resolución de problemas, así como el desarrollo continuo de la plataforma.

Algunas de las instituciones que han utilizado *CollectiveAccess* para publicar sus colecciones digitales son el *Queens Memory Project*⁴⁴, el *New Museum's Digital Archive*⁴⁵, el *Parrish Eastend Stories*⁴⁶, el *Van Alen Institute's Design Archive*⁴⁷, el repositorio del *Jewish Museum*⁴⁸ o, dentro de nuestras fronteras, la *Kutxateka*⁴⁹.

3.3. Greenstone

Se trata de un conjunto de programas de código abierto y multiplataforma, diseñados para crear y distribuir colecciones digitales a través de Internet o en formato CD-ROM. El proyecto fue iniciado en la Universidad de Waikato, Nueva Zelanda, a principios de la década de 2000. Desde entonces, ha ganado popularidad en todo el mundo y ha sido adoptado por muchas bibliotecas, museos, archivos y otras instituciones que desean crear y distribuir colecciones digitales de una manera organizada y accesible. Su comunidad de usuarios y desarrolladores es activa y colaborativa, lo que permite obtener soporte y compartir recursos con otros usuarios en todo el mundo.

Una de las características más destacadas de *Greenstone* es su capacidad para crear colecciones digitales interactivas y personalizadas. Los usuarios pueden cargar una variedad de tipos de contenido, como texto, imágenes, audio y video, y organizarlos en colecciones temáticas. Esto es especialmente útil para bibliotecas y archivos que desean proporcionar acceso a materiales digitales diversificados, como libros, revistas, fotografías históricas, grabaciones de audio y videos educativos.

Greenstone también facilita la búsqueda y navegación de los contenidos de la colección. Los usuarios pueden realizar búsquedas avanzadas utilizando palabras clave, metadatos y otros

⁴⁴ <http://queensmemory.org/>

⁴⁵ <http://archive.newmuseum.org/>

⁴⁶ <http://artists.parrishart.org/>

⁴⁷ <http://archive.vanalen.org/archive/>

⁴⁸ <http://collections.jewishmuseum.cz/>

⁴⁹ <https://www.kutxateka.com/>

criterios de búsqueda. Además, las colecciones pueden ser navegadas de manera intuitiva a través de una interfaz de usuario amigable que incluye opciones de filtrado y categorización. Esto hace que sea más fácil para los usuarios encontrar y acceder a la información que están buscando.

Otra característica importante de *Greenstone* es su capacidad para admitir múltiples idiomas y scripts, lo cual resulta esencial para instituciones que desean ofrecer sus colecciones digitales a audiencias internacionales y que pueden tener necesidades específicas de idioma.

Además de su funcionalidad básica, ofrece una serie de características avanzadas que pueden mejorar la experiencia del usuario y la gestión de las colecciones digitales. Estas incluyen la capacidad de crear presentaciones multimedia interactivas, integración con sistemas de gestión de contenidos (CMS) y la posibilidad de personalizar la apariencia y la estructura de la interfaz de usuario.

Es importante destacar que, aunque *Greenstone* ha sido utilizado en una variedad de proyectos, muchos de ellos ya no están accesibles o han migrado a otras plataformas. Entre los que se encuentran disponibles se pueden mencionar ARCAS, de la Universidad Nacional de La Plata (Argentina)⁵⁰, *Armenian Rare Books*⁵¹ o *Manioc*⁵², una biblioteca digital desarrollada por la Universidad de las Antillas y Guyana.

3.4. CollectionBuilder

CollectionBuilder es un *framework* de código abierto diseñado para ofrecer una forma sencilla de publicar colecciones digitales en línea, que no involucre demasiados conocimientos informáticos ni infraestructura técnica. Desarrollado en 2019 por la biblioteca de la Universidad de Idaho y el *Center for Digital Inquiry and Learning* (CDIL), sigue la metodología Lib-STATIC.

El concepto *framework*, se refiere al hecho de que *CollectionBuilder* no es un software en sí mismo, sino una plantilla que al compilarse da como resultado un sitio web estático completamente funcional. Dicha plantilla, no es más que una carpeta estructurada en la que se alojan los objetos digitales, un CSV con sus metadatos, archivos de configuración y plantillas de página basadas en HTML, CSS y *JavaScript*⁵³ (*Figura 1*).

⁵⁰ <http://arcas.fahce.unlp.edu.ar/>

⁵¹ <http://greenstone.flib.sci.am/gsd/cgi-bin/library.cgi?e=p-00000-00---off-0--00-----0-10-0---0---odirect-10---4-----0-1--10-en-50---20-home---0--1-00-0-0-01-1-outfZz-8-00&a=p&p=about&c=armenian>

⁵² <http://www.manioc.org/>

⁵³ HTML, CSS y *JavaScript* son tres tecnologías fundamentales en el desarrollo de páginas web. HTML se utiliza para estructurar y organizar el contenido de la página, definiendo elementos como encabezados, párrafos, imágenes y enlaces. CSS complementa HTML al permitir la definición del estilo y diseño de estos

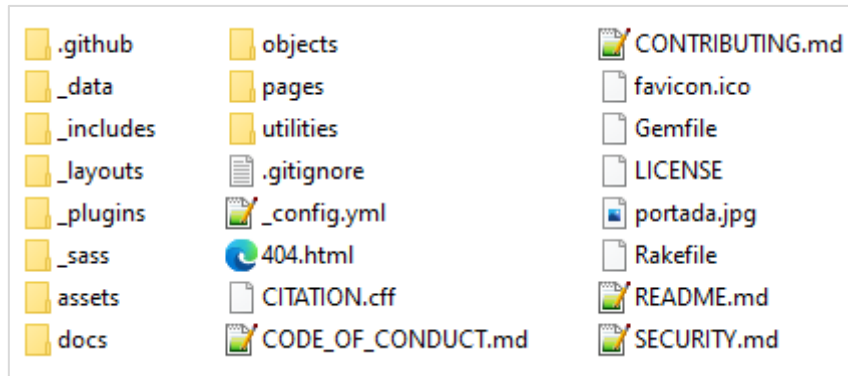


Figura 1- Estructura de la plantilla CollectionBuilder-CSV

Esta estructura organizativa permite crear exposiciones digitales atractivas sin tener que invertir mucho esfuerzo. El *framework* utiliza tecnologías web modernas, como *Jekyll* (un generador de sitios estáticos), que permite generar un sitio web rápido, ligero y de fácil mantenimiento, a partir de la plantilla y los diseños predefinidos (ahorrando mucho tiempo y esfuerzo en la construcción de la estructura básica del sitio), y que puede ser personalizado según las necesidades de la colección. Esto permite que cada exposición tenga su propia identidad visual y se ajuste a su contenido específico.

El *framework* también es altamente flexible en términos de escalabilidad, ya que puede utilizarse tanto para crear exposiciones digitales simples, con unas pocas docenas de objetos, como para proyectos más grandes que involucren cientos o miles de ellos.

CollectionBuilder cuenta con dos versiones, *CollectionBuilder-GH* y *CollectionBuilder-CSV*.

- **CollectionBuilder-GH:** es una plantilla diseñada específicamente para ser alojada en *GitHub Pages*, por lo que permite publicar y editar las colecciones sin necesidad de instalar nada. Es por tanto la opción ideal para iniciarse en la utilización de este *framework*. No obstante, está pensado para colecciones pequeñas, ya que la versión gratuita de *GitHub* solo ofrece 1GB de almacenamiento.
- **CollectionBuilder-CSV:** se trata de una plantilla que se descarga y edita desde el ordenador para luego poder publicarla en un servidor, ya sea local o de Internet.

En ambos casos, el tema se basa en *Bootstrap*⁵⁴, mientras que los distintos tipos de visualizaciones de datos que ofrecen (líneas de tiempo, mapas, galerías, etc.) se crean utilizando bibliotecas

elementos, mientras que *JavaScript* es un lenguaje de programación que permite agregar interactividad y dinamismo al sitio.

⁵⁴ *Bootstrap* es una colección de herramientas basadas en CSS y *JavaScript* orientadas a facilitar la creación de sitios y aplicaciones web. Ofrece estilos y elementos prediseñados que se pueden utilizar y adaptar según las necesidades, lo que simplifica el proceso de diseño y ahorra tiempo y esfuerzo en la creación de interfaces atractivas y funcionales.

JavaScript de código abierto como *DataTables*, *Leafletjs*, *Spotlight gallery*, *lazysizes* y *Lunr.js*⁵⁵. Asimismo, los metadatos de los objetos se exponen utilizando los estándares de protocolo *Schema.org* y *Open Graph*.

CollectionBuilder-GH vs CollectionBuilder-CSV

Si bien el funcionamiento de ambas opciones es similar, no cabe duda de que *CollectionBuilder-GH*, representa la opción más sencilla de implementar, en tanto es una solución que no requiere de la administración de un *hosting* propio; no obstante, esta misma ventaja se convierte en su principal inconveniente cuando la colección que se desea publicar crece por encima del límite de almacenamiento ofrecido de forma gratuita. Es ahí, cuando *CollectionBuilder-CSV* se antoja como la única opción a valorar.

Por otra parte, como ya se ha comentado, la generación del sitio web resultante se lleva a cabo empleando el constructor de sitios web estáticos *Jekyll*. En el caso de la plantilla *CollectionBuilder-GH*, este proceso se realiza de forma transparente mediante un flujo de trabajo configurado previamente en *GitHub Actions*⁵⁶. Por el contrario, la versión *CollectionBuilder-CSV* requiere tener instalado previamente *Ruby*, además de las *gemas*⁵⁷ *Jekyll* y *Bundler*⁵⁸, así como todas las dependencias necesarias (y que se encuentran especificadas en el archivo *Gemfile.lock* incluido en la plantilla). Cabe señalar que, tanto la instalación de las gemas como de las dependencias se debe llevar a cabo a través de un intérprete de comandos. Del mismo modo, la ejecución de *Jekyll* para compilar la colección también se debe llevar a cabo mediante comandos.

Dado que *CollectionBuilder* es un proyecto relativamente joven en el ámbito de las herramientas

⁵⁵ Cada una de estas herramientas desempeña un papel crucial en la creación de experiencias web ricas y funcionales. *DataTables* simplifica la visualización y manipulación de datos tabulares en páginas web, permitiendo búsqueda, filtrado y paginación dinámicos, *Leaflet.js* es una biblioteca que facilita la incorporación de mapas personalizados en sitios web, con soporte para geolocalización y marcadores interactivos, *Spotlight gallery* permite la creación de galerías de imágenes y medios enriquecidos de manera atractiva y altamente personalizable, *Lazysizes* optimiza la carga de imágenes y otros recursos multimedia, mejorando significativamente el rendimiento de las páginas web al cargar contenido de manera diferida, mientras que *Lunr.js* es una biblioteca que permite añadir capacidades de búsqueda a texto completo de forma rápida y eficiente, sin necesidad de depender de motores de búsqueda externos.

⁵⁶ Las *GitHub Actions* son una característica de *GitHub*, una plataforma de desarrollo colaborativo, que permite automatizar tareas y flujos de trabajo que se ejecutan en respuesta a eventos específicos.

⁵⁷ En el lenguaje de programación *Ruby*, una "gema" (*gem*, en inglés) es una unidad de software empaquetada que contiene código reutilizable y otros recursos necesarios para obtener una funcionalidad específica. Las gemas son una parte fundamental del ecosistema de *Ruby* y permiten a los desarrolladores compartir, distribuir y reutilizar código de manera eficiente.

⁵⁸ *Bundler* es una herramienta de *Ruby* que administra las dependencias de un proyecto. Garantiza que todas las gemas necesarias se instalen y funcionen de manera consistente, evitando problemas de compatibilidad y facilitando la colaboración entre desarrolladores.

de creación de colecciones digitales, no existen muchos ejemplos de instituciones que lo hayan implementado en comparación con otras soluciones más establecidas. En su mayoría, los ejemplos de implementación de *CollectionBuilder* se han creado en el contexto de actividades formativas y de capacitación, donde estudiantes y bibliotecarios han utilizado la plataforma como una herramienta de aprendizaje para desarrollar exhibiciones y colecciones digitales⁵⁹, lo que demuestra su creciente interés y potencial en el ámbito educativo y cultural. A medida que el proyecto madure, es posible que más instituciones comiencen a adoptar *CollectionBuilder* para sus proyectos de colecciones digitales.

3.5. Comparativa

Para recoger las características de todas las aplicaciones analizadas, se ha consultado su documentación oficial y se han instalado en un entorno local o de servidor o, bien se han consultado sus versiones de demostración, según el caso. La siguiente tabla recoge los resultados de la comparativa.

	<i>Omeka S</i>	<i>CollectiveAccess</i>	<i>Greenstone</i>	<i>CollectionBuilder CSV</i>
Sitio web oficial	https://omeka.org/s	https://www.CollectiveAccess.org/	https://www.greenstone.org/index_es	https://CollectionBuilder.github.io/
Año de creación	2017	2007	2000	2019
Última versión estable	4.0.4	1.7.16	3.11 (Greenstone3)	N/A
Desarrollador	Departamento de Historia e Historia del Arte de la Universidad George Mason	Whirl-i-Gig en colaboración con entidades como el Institute of Museum and Library Services	Universidad de Waikato en colaboración con la UNESCO y la ONG de <i>Human Info</i>	Biblioteca de la Universidad de Idaho y el Center for Digital Inquiry and Learning
Licencia	GPLv3	GPLv3	GPLv3	MIT license
Lenguaje de programación	PHP	PHP	Java	Ruby
Gestor base de datos	MySQL o MariaDB	MySQL	MGDB (propio). Admite MySQL, PostgreSQL y SQLite	N/A
Soporte servidor web	Apache o Nginx	Apache o Nginx	Apache o Nginx	N/A

⁵⁹ Sirva como ejemplo *Retroproyector* (<https://cdigitaljaveriana.github.io/collectionbuilder-gh/>), que recoge artefactos y materiales pre-digitales almacenados en la Biblioteca Alfonso Borrero Cabal. Este proyecto hace parte de los ejercicios realizados por estudiantes del curso Cultura Digital en la Pontificia Universidad Javeriana de Bogotá.

Formatos de importación	CSV	Excel, Google Spreadsheets, CSV, marc21, FilemakerPro, DSO XML e InMagick XML.	CSV, XML, EXIF, MARC, CDS/ISIS, ProCite, BibTex, Refer, OAI, DSpace y METS	CSV
Formatos de exportación	CSV	XML, MARC21 y CSV	XML y CSV	CSV, XML y JSON
Metadatos	Se puede integrar cualquier vocabulario en formato ontología.	Se puede integrar cualquier esquema de metadatos en XML.	Dublin Core (calificado y no calificado) RFC 1807 NZGLS (Servicio de localización del gobierno de Nueva Zelanda) AGLS (Servicio de localización del gobierno australiano)	Dublin Core
Interoperabilidad OAI-PMH	Sí	Sí	Sí	Sí
Motor de búsqueda	MySQL y Solr	MySQL y Solr	Solr	Lunr.js
Búsqueda avanzada	Sí	Sí	Sí	Sí

Tabla 1 - Comparativa entre Omeka S, CollectiveAccess, Greenstone y CollectionBuilder-CSV.

Fuente: elaboración propia.

Conclusiones

Sin lugar a duda, tanto *Omeka S* como *CollectiveAccess* representan opciones muy completas en el ámbito de la gestión de repositorios y bibliotecas digitales. Sin embargo, es importante destacar que ambas aplicaciones requieren una inversión inicial más significativa en la configuración de la arquitectura del servidor web necesario, en contraste con las alternativas restantes.

Al adentrarnos en las diferencias entre *Omeka S* y *CollectiveAccess*, podemos apreciar que este último se distingue por su relativa facilidad para integrar diversos esquemas de metadatos en el sistema. Esto se debe, en parte, a la disponibilidad de archivos de configuración para los principales lenguajes, lo cual agiliza la incorporación de estándares específicos. Además, *CollectiveAccess* sobresale en la gestión de usuarios y permite una personalización más completa de la aplicación para adaptarla a las necesidades específicas del proyecto. En contraposición, *Omeka S* presenta una curva de aprendizaje menos pronunciada para los administradores, lo que facilita su adopción inicial. Además, ofrece un conjunto más amplio de herramientas para crear nuevos productos documentales, como exposiciones virtuales, líneas de tiempo o mapas. Estas herramientas se pueden extender aún más gracias a la disponibilidad de módulos que simplifican su instalación y personalización.

En lo que respecta a *Greenstone* y *CollectionBuilder*, también se pueden identificar ventajas y características únicas. *Greenstone*, con una historia que se remonta a los años dos mil, brinda una flexibilidad significativa en la gestión de metadatos y admite una amplia variedad de formatos de importación y exportación, pero su interfaz predefinida se siente añeja. Por otro lado, *CollectionBuilder*, al ser un sistema mucho más reciente ofrece una interfaz más atractiva visualmente y una experiencia de usuario más ágil, y resulta ideal para proyectos simples, basados en metadatos en formato CSV.

En resumen, la elección de una u otra solución dependerá en última instancia de los objetivos y necesidades específicas del proyecto. Tanto *Omeka S* como *CollectiveAccess* requieren un esfuerzo significativo en su configuración inicial, aunque difieren en lo que respecta a la integración de metadatos y la facilidad de personalización. *Greenstone* ofrece una gran flexibilidad en la gestión de metadatos y formatos, pero actualmente su uso es testimonial, mientras que *CollectionBuilder* se adapta mejor a proyectos más sencillos a costa de sacrificar funcionalidades avanzadas.

4. Marco de trabajo propuesto

Como se ha visto, el proceso de crear y publicar una colección digital debe comenzar con una estrategia que esté alineada perfectamente con la institución que lo impulsará. Esta necesidad implica realizar una cuidadosa planificación, así como el diseño de un marco de trabajo que abarque la descripción, la serialización y la elección de las aplicaciones informáticas que lo sustentarán. Además, es de suma importancia considerar todo el ciclo de vida de la colección digital y los servicios relacionados que la acompañan, así como garantizar la accesibilidad futura y el mantenimiento continuo de las colecciones.

En este contexto se plantea la cuestión de si existe alguna solución que en su conjunto responda a principios de computación mínima, esto es, que requiera los menores recursos posibles de software, hardware, conocimientos informáticos, etc. y que pueda ser aplicada en una amplia variedad de casos.

4.1. Especificaciones del producto final

El primer paso para poder responder a esta pregunta consistió en definir detalladamente las características que debía cumplir el marco de trabajo para ajustarse a estos requerimientos.

De este modo se estableció que el producto resultante debía ser un espacio digital, perfectamente localizado a idioma español, que respondiera al reto de describir y divulgar el acervo documental de cualquier tipo de institución de la memoria, a fin de facilitar su consulta y promover la investigación, y en el cual se pudieran encontrar los objetos digitales clasificados en categorías temáticas y tipos documentales. De igual manera, de cada objeto se podría obtener información como fechas, datos de ubicación, una descripción detallada del contenido, temas, referencias de material complementario disponible, derechos de uso, etc.

Por otra parte, dado que cualquier sistema de información, ya sea en el mundo tangible o en el ciberespacio, necesitan establecer el modelo de base de datos que emplearán para llevar a cabo el tratamiento documental de los objetos con los que cuentan, fue necesario definir también sus características:

- Estar basada en software libre: esto implica que la base de datos debe utilizar software cuyo código fuente es accesible y modificable por cualquiera. Esta característica fomenta la transparencia, la colaboración y la independencia de proveedores, además de reducir los costes de licencia.
- Ser ligera: la ligereza se refiere a la eficiencia en el uso de recursos del sistema, como memoria y CPU. Una base de datos ligera se ejecuta de manera eficiente en sistemas con recursos limitados, lo que es fundamental para garantizar un rendimiento óptimo en una variedad de entornos.
- Fácil de implementar: la instalación y configuración de la base de datos serán simples y sin complicaciones. Esto reduce el tiempo y los recursos requeridos para poner en

funcionamiento el sistema, lo que es especialmente valioso en proyectos con plazos ajustados o personal con bajos conocimientos técnicos.

- Requerir los menores recursos de hardware posible: esto se relaciona con la eficiencia mencionada anteriormente. Una base de datos que requiere pocos recursos de hardware garantiza que pueda funcionar en ordenadores menos potentes o en servidores compartidos, lo cual ahorra costes y facilita su escalabilidad.
- Estar guiada por pautas internacionales en el ámbito de la catalogación patrimonial: esta característica implica que la estructura y el contenido de la base de datos seguirán estándares y directrices internacionalmente reconocidos en la catalogación de patrimonio, asegurando la coherencia y la calidad de los datos almacenados.
- Ser compatible con estándares de interoperabilidad: la interoperabilidad se refiere a la capacidad de la base de datos para comunicarse y compartir datos con otros sistemas y aplicaciones. Ser compatible con estándares de interoperabilidad garantiza que la base de datos pueda integrarse de manera efectiva en un ecosistema tecnológico más amplio.
- Ofrecer diferentes niveles de búsqueda según el perfil del usuario y su exigencia de información; esto significa que la base de datos proporcionará una variedad de opciones de búsqueda que se adaptarán a las necesidades de diferentes usuarios. Desde búsquedas simples y rápidas hasta consultas más avanzadas, los usuarios podrán acceder a la información de manera eficiente y personalizada según sus requerimientos.

En conjunto, estas características aseguran que la base de datos sea una herramienta versátil y eficaz para la gestión y búsqueda de información patrimonial, al mismo tiempo que cumple con los principios de accesibilidad, eficiencia y calidad de datos.

Otra cuestión esencial para garantizar una experiencia positiva para los visitantes es la interfaz del sitio web. Además de la modernidad, la simplicidad y la estética son factores cruciales que se deben considerar al diseñar plataformas en línea.

Para lograr una apariencia moderna, es importante adoptar un diseño limpio y minimalista. Esto implica prescindir de elementos innecesarios que puedan distraer o abrumar a los usuarios. La simplicidad no solo se refiere a la cantidad de elementos en la página, sino también a la disposición y estructura de la información. Un diseño intuitivo y fácil de navegar facilita la exploración de la colección y garantiza que los visitantes puedan encontrar rápidamente lo que están buscando.

La estética es otro aspecto crucial. La selección de colores, tipografía y elementos visuales debe ser coherente con la identidad de la colección y atractiva para el público objetivo. Los colores pueden transmitir emociones y conceptos, por lo que elegir una paleta que refleje la esencia de la colección es esencial. Asimismo, la tipografía debe ser legible y complementar el estilo general del sitio.

Además de la apariencia visual, la funcionalidad también es fundamental. El sitio web debe ser responsivo, es decir, adaptarse a diferentes dispositivos y tamaños de pantalla para garantizar una experiencia óptima en smartphones, tabletas y ordenadores. También es importante garantizar una carga rápida de las páginas para no frustrar a los visitantes con tiempos de espera prolongados.

En resumen, la interfaz del sitio web de la colección debe combinar modernidad, simplicidad y estética para brindar a los usuarios una experiencia agradable y efectiva. Esto no solo atraerá a un público más amplio, sino que también garantizará que la colección se presente de la mejor manera posible en el entorno digital.

4.2. Componentes

En base a todas estas características y a las conclusiones obtenidas tras el análisis del estado de la cuestión, se hace evidente que, de todas las soluciones estudiadas, *CollectionBuilder* es la que más se adecua al concepto de computación mínima, ya que se desmarca claramente del resto por no requerir un entorno AMP (Apache, MySQL y PHP)⁶⁰ para su funcionamiento. Esto significa que no depende de una infraestructura de servidor compleja ni de bases de datos en tiempo real, lo cual tiene un impacto significativo en la eficiencia de la implementación, ya que reduce la necesidad de recursos de hardware y el consumo de ancho de banda de Internet en comparación con las demás alternativas estudiadas.

Asimismo, cumple el requisito de adherirse a pautas internacionales reconocidas en la catalogación de patrimonio cultural, como *Dublin Core*, lo cual garantiza que los datos de la colección se organicen y estandaricen de manera coherente y uniforme, cuestión esencial para asegurar la integridad y la calidad de la información almacenada. También ofrece a los usuarios múltiples niveles de búsqueda, permitiendo encontrar información de manera rápida y sencilla, o realizar búsquedas más detalladas para obtener resultados precisos.

Otra cuestión en la que destaca *CollectionBuilder*, es en la adaptabilidad. Su visualización se ajusta perfectamente tanto a tabletas como a dispositivos móviles, lo cual es esencial para garantizar una amplia difusión de las colecciones digitales. Esta adaptabilidad se logra a través de tecnologías web modernas y eficientes, posibilitando una experiencia de usuario ágil y atractiva.

Es lógico, por tanto, considerar que *CollectionBuilder* supone una buena alternativa en aquellos casos en los que no se cuenta con recursos informáticos avanzados, que no se disponga del tiempo necesario para aprender a utilizar algo más complejo y, en resumidas cuentas, cuando se quiere reutilizar de forma rápida colecciones digitalizadas para brindar acceso a ellas a la comunidad.

⁶⁰ Es habitual que *Apache*, *MySQL* y *PHP* colaboren de manera conjunta en el desarrollo de un sitio web, ya que cada uno desempeña un rol fundamental en este proceso. *Apache* se encarga de servir las páginas web, *PHP* procesa la lógica del lado del servidor y *MySQL* administra la base de datos que almacena la información.

Sin embargo, aunque es una herramienta valiosa para la creación de exposiciones digitales, también presenta inconvenientes. Algunos usuarios pueden encontrar complicada su puesta en funcionamiento inicial, especialmente si no están familiarizados con los intérpretes de comandos y con las tecnologías web, o, por el contrario, considerar sus funcionalidades por defecto demasiado básicas, que, si bien pueden ser subsanadas, requeriría de conocimientos de desarrollo web avanzados.

Dejando a un lado estos hándicaps, no cabe duda de que la filosofía que subyace del proyecto se alinea perfectamente con los objetivos de esta investigación, y dado que su carácter de código abierto posibilita su adaptación a las especificaciones propuestas, se decide desarrollar el marco de trabajo a partir de *CollectionBuilderCSV*, con el objetivo de simplificar más aún si cabe el proceso de publicación de una colección digital. De este modo, tras el estudio detallado de su funcionamiento se determina que el marco estará formado por tres elementos distintos:

- una plantilla del sitio web,
- una plantilla de metadatos y
- un software que se encargará de “hilar” automáticamente ambos componentes con los correspondientes objetos digitales.

4.2.1. Plantilla del sitio web

La plantilla del sitio web fue construida mediante un *fork* de la plantilla original de *CollectionBuilder CSV*, clonando su correspondiente repositorio de *GitHub* y publicándola en <https://github.com/lpitac/exposicion>.

La primera modificación introducida fue localizarla completamente a español, lo cual se llevó a cabo de forma manual, partiendo de las indicaciones básicas que se proveen en la documentación oficial.

Seguidamente se modificó la página principal, añadiendo un buscador avanzado y reorganizando las secciones. También se corrigió el código del carrusel de imágenes para que se adaptase al tamaño de pantalla disponible, ya que en el desarrollo original el escalado no se realizaba de forma coherente.

En lo que respecta a la apariencia general del sitio, se modificó el color principal del tema y se realizaron algunos cambios menores relativos a la cabecera y al pie de la página, al tamaño y formato de las tipografías, etc.

Aunque *CollectionBuilder* ofrece una plantilla de página para generar visualizaciones de los datos sobre un mapa (basándose en las coordenadas establecidas en sus metadatos), se decidió prescindir de esta opción ya que su utilidad se antoja residual y su omisión contribuye a aligerar el peso del desarrollo final. En esta misma línea también se trataron de purgar todos aquellos archivos innecesarios relativos a librerías y plantillas de páginas no utilizadas, dando como resultado que la plantilla final pese algo menos de 20MB .

4.2.2. Plantilla de metadatos

El segundo componente del marco de trabajo propuesto se desarrolló a partir de la propia plantilla de metadatos de *CollectionBuilder-CSV*. Esta plantilla se basa en una hoja de *Google Sheets*⁶¹ que, una vez editada, se puede descargar en formato CSV. Se encuentra compartida en https://docs.google.com/spreadsheets/d/1j330xd6E3-PKpFYi_oTGbh8sARu-VMDdkZ5w9datpyo/edit#gid=835025159.

Nótese que la elección de *Google Sheets* como plataforma de edición no fue una decisión tomada al azar, sino que está basada en consideraciones técnicas y prácticas. En primer lugar, porque herramientas como *Microsoft Excel* pueden ocasionar problemas en la exportación a CSV, especialmente cuando se intentan utilizar opciones avanzadas como listas desplegables, fórmulas y otras funcionalidades. Además, *Excel* tiende a distorsionar cualquier carácter que no sea ASCII, lo que podría comprometer la integridad de los datos.

Por otro lado, la edición de la plantilla a través de *Google Sheets* ofrece varias ventajas significativas. Permite aprovechar todas las funcionalidades avanzadas de la plataforma, lo que facilita la introducción de datos y el control de errores, y, por tanto, la optimización y validación de los datos antes de su conversión a formato CSV.

En este sentido, las principales mejoras introducidas en la plantilla desarrollada con respecto a la original consistieron en crear reglas de validación basadas en menús desplegables para algunos datos, así como campos que se “autocalculan” gracias a fórmulas preestablecidas. Asimismo, también se trató de darle un aspecto más limpio, ordenado y visualmente atractivo que la propuesta original.

Cabe reseñar que esta plantilla también se trató de localizar a idioma español por completo, si bien los campos relacionados con ciertos metadatos se mantuvieron con su nombre original en inglés, ya que su traducción requeriría de modificaciones a nivel de librerías, lo cual se consideró inabarcable e innecesario.

El esquema de metadatos se basa en *Dublin Core* y está formado por 18 atributos, cada uno de los cuales desempeña un papel crucial en la catalogación y descripción de los objetos, proporcionando información esencial sobre su origen, contenido, contexto temporal y geográfico, entre otras cuestiones.

A continuación, se ofrece una relación detallada de cada uno de los metadatos contemplados en el esquema diseñado:

- "objectid" sirve como identificador único para cada uno de los documentos en el *dataset*.

⁶¹ *Google Sheets* es una aplicación de hojas de cálculo en línea que permite crear y editar documentos de manera colaborativa en tiempo real. Ofrece funciones avanzadas y almacenamiento en la nube para una gestión de datos eficiente y accesible desde cualquier dispositivo.

- "nombre de archivo" contiene el nombre del archivo asociado a cada documento.
- "parentid" permite establecer relaciones entre los objetos.
- "display_template" clasifica y define la categoría o tipo de cada documento dentro del conjunto documental.
- "object_location" proporciona información relativa a la ubicación del objeto en el repositorio.
- Las URLs "image_small" e "image_thumb" permiten acceder a imágenes en tamaños pequeños y miniaturas, facilitando la visualización de los documentos.
- "titulo" registra el título o nombre de cada uno de los documentos.
- "autor" identifica al autor o creador del documento.
- "fecha" asigna una fecha asociada al documento, brindando contexto temporal.
- "date-is-approximate?" indica si la fecha es una aproximación.
- "descripcion" brinda una breve descripción del contenido de cada documento.
- "tema" identifica temas relacionados con cada documento.
- "idioma" especifica el idioma en el que está redactado el contenido de cada documento.
- "fuente" establece la procedencia de los documentos, destacando su origen.
- "licencia" indica la licencia bajo la cual se comparten los documentos.
- "localizacion" contextualiza la ubicación geográfica o el contexto relacionado con cada documento.
- "image_alt_text" proporciona un texto alternativo para las imágenes asociadas a los documentos.
- "object_transcript" incluye la transcripción o el contenido textual de cada documento.

Una cuestión a señalar es que este esquema aborda la inclusión de objetos compuestos dentro del conjunto de datos mediante el campo "parentid". Un "objeto compuesto" se define aquí como un elemento que consta de un conjunto de archivos digitales destinados a ser tratados como una entidad conectada dentro de la exposición virtual y que se presenta en una sola página de elemento. Estos objetos compuestos pueden abarcar diversos tipos de archivos, incluyendo imágenes, PDF, vídeos y audio. Un ejemplo ilustrativo lo representan las postales, que típicamente están compuestas por el anverso y el reverso. Sin embargo, esta concepción de objetos compuestos es igualmente aplicable a cualquier compendio que se desee crear. Este enfoque del esquema de metadatos garantiza que se pueda capturar y describir de manera efectiva la complejidad de este tipo de objetos, permitiendo así la gestión y consulta eficaz de sus relaciones dentro del conjunto de datos.

4.2.3. Constructor automático de la colección

Como ya se comentó, el proceso de creación y actualización de una colección digital mediante *CollectionBuilder CSV* requiere tener instalado *Ruby* y hacer uso de un intérprete de comandos, lo cual presenta varios inconvenientes significativos; no todos los usuarios, especialmente aquellos que no tienen experiencia en programación o desarrollo web, están familiarizados con

este lenguaje o se sienten cómodos utilizando un intérprete de comandos. Esto crea una barrera técnica para muchas personas que pueden estar interesadas en crear o mantener una colección digital. La necesidad de instalar *Ruby* y aprender a usarlo puede resultar intimidante y desalentador.

Cabe tener en cuenta, además, que *Ruby* es un lenguaje de programación que funciona mejor en sistemas basados en *Unix* o *Linux*. Esto significa que los usuarios que tienen sistemas operativos diferentes, como *Windows*, pueden enfrentarse a dificultades adicionales al intentar configurar y usar *Ruby* con *CollectionBuilder-CSV*. Esto reduce aún más el acceso y la usabilidad para un público más amplio. Además, para quienes no tienen experiencia previa con este lenguaje o con la línea de comandos, la curva de aprendizaje puede resultar muy pronunciada, lo cual puede alejar a muchos usuarios potenciales que prefieren una solución más amigable y orientada a la interfaz gráfica. A ello hay que añadir que el uso de la línea de comandos y scripts personalizados conlleva el riesgo de cometer errores técnicos que pueden ser difíciles de diagnosticar y solucionar, provocando que se pierdan datos y tiempo en la corrección de errores.

Por otra parte, *Ruby* y sus dependencias, así como las tecnologías utilizadas para desarrollar *CollectionBuilder*, pueden requerir actualizaciones periódicas para mantenerse al día con cambios en la tecnología o para solucionar problemas de seguridad. Esto agrega una responsabilidad continua de mantenimiento que no todos los usuarios pueden estar dispuestos o capacitados para asumir.

En resumen, la dependencia de *Ruby* y la necesidad de utilizar la línea de comandos para crear y mantener una colección digital basada en *CollectionBuilder-CSV* limita significativamente la accesibilidad y la usabilidad de esta herramienta para un público más amplio. Esto puede ser un gran inconveniente, ya que restringe el acceso a personas que podrían estar interesadas en compartir su contenido en línea pero que no tienen la experiencia requerida para hacerlo de manera efectiva.

Fiador

Para superar esta barrera técnica y hacer que la creación de colecciones digitales sea más accesible, se escribió un software en *Python* que eliminase dichos hándicaps. Este software se basa en un desarrollo propio, que ha sido bautizado con el nombre de “Fiador”, y que se centra en simplificar y automatizar el proceso de vincular objetos digitales con sus metadatos en formato CSV, para posteriormente generar automáticamente todos los archivos necesarios para dar forma al sitio web resultante, en base a la plantilla desarrollada previamente. De este modo, en lugar de depender de la línea de comandos, *Fiador* presenta una interfaz gráfica de usuario intuitiva y amigable, fácil de usar incluso para aquellos que no tienen experiencia técnica.

En lo que respecta a la elección de *Python* como lenguaje de programación, cabe reseñar que se trató de una decisión estratégica justificada por varias razones. *Python* es ampliamente reconocido por su simplicidad y facilidad de aprendizaje. Su sintaxis es clara y legible, lo cual lo convierte en un lenguaje ideal para personas sin experiencia previa en programación. Cuenta con

una comunidad de desarrolladores extremadamente activa y una gran cantidad de recursos disponibles en línea. Esto implica que los usuarios que enfrenten dudas o desafíos pueden encontrar fácilmente documentación, tutoriales y foros de discusión que les brinden la ayuda necesaria. Tiene, además, una gran cantidad de bibliotecas y módulos que facilitan el desarrollo de software personalizado. Esto es particularmente útil para poder diseñar una interfaz gráfica de usuario (GUI) intuitiva o manejar operaciones de lectura y escritura de archivos CSV, como se requiere en este contexto. Por otro lado, al ser un lenguaje de programación de código abierto, encaja perfectamente con la filosofía de este proyecto. También resulta muy versátil y se utiliza en una amplia variedad de aplicaciones, desde desarrollo web hasta análisis de datos y automatización. Esto significa que, en el futuro, *Fiador* podría ampliarse o modificarse para incluir características adicionales o adaptarse a diferentes necesidades de los usuarios.

A todo ello hay que añadir que se trata de un lenguaje multiplataforma, lo que significa que las aplicaciones desarrolladas pueden ejecutarse en una variedad de sistemas operativos, incluidos *Windows*, *Linux* y *MacOS*, si bien requiere tener instalado *Python* previamente. Esto enseguida se reveló como un hándicap para lograr la facilidad de uso del programa, así como la utilización del mínimo software posible, por lo que se optó por *paquetizar* el código resultante en un ejecutable *.exe*, aún a riesgo de restringir su uso únicamente a sistemas *Windows*.

Tras varias versiones de desarrollo, originadas como consecuencia de los múltiples desafíos y errores que se iban presentando, se consiguió obtener un “producto mínimo viable”⁶² que sirviese como prueba de concepto para los objetivos de esta investigación, y que cubriese todos los procesos implicados en la generación y edición de los elementos básicos del sitio web de la colección.

4.3 Flujo de tareas

En el contexto de la publicación de colecciones digitales, la implementación del marco de trabajo propuesto implica seguir el flujo de tareas que a continuación se detalla, y que se encuentra compuesto por tres sencillos pasos:

El primero consiste en realizar una copia de la plantilla de metadatos diseñada y alojada en *Google Sheets*. Esto se puede hacer descargándola localmente en formato CSV o, mejor aún, clonándola directamente en otra instancia de *Sheets* a la que se tenga acceso, para así poder aprovechar las facilidades de edición que proporciona. De este modo, una vez clonada, simplemente será necesario editarla para incluir los metadatos propios de la colección a describir, y, posteriormente, proceder a su descarga local en formato CSV.

El segundo paso consiste en reunir en una misma carpeta todos y cada uno de los objetos digitales descritos en el fichero CSV, y que serán relacionados por el campo “nombre de archivo”.

⁶² *Fiador 1.0* se encuentra publicado en <https://github.com/lpitac/fiador/tree/main> bajo licencia GNU General Public License 3.0 (<https://www.gnu.org/licenses/gpl-3.0.html>)

Por último, se debe descargar *Fiador* desde su repositorio oficial y proceder a su ejecución. En la ventana que se abrirá se mostrarán las tres opciones disponibles “Generar exposición”, “Editar exposición” y “Abrir exposición”.



Figura 2 - Ventana principal de Fiador 1.0 beta

4.3.1. Generar una exposición

Hay dos tipos de exposición que se pueden crear:

1. **Exposición de prueba:** utiliza los objetos digitales y el CSV con sus correspondientes metadatos incluidos en la plantilla.
2. **Exposición personalizada:** permite definir los objetos digitales y el CSV que se utilizarán para generar el sitio web.

En ambos casos se podrán editar los datos a posteriori.



Figura 3 - Fiador 1.0. Ventana "Generar exposición"

Exposición de prueba

Para generar una exposición de prueba tan solo será necesario indicar la carpeta local en la que se desea albergar y pulsar sobre el botón “Generar”.

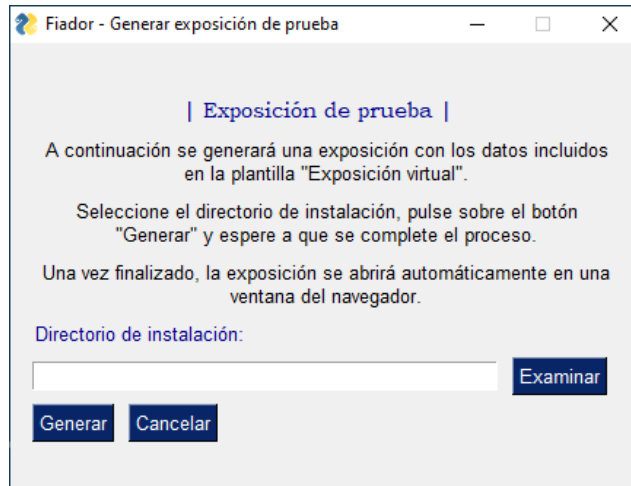


Figura 4 - Fiador 1.0. Ventana "Exposición de prueba"

El proceso tardará varios minutos en completarse. Una vez listo, se mostrará el mensaje "¡Exposición generada correctamente!".

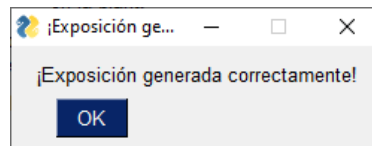


Figura 5 - Fiador 1.0. "Pop-up" exposición generada correctamente.

Al pulsar sobre el botón "Ok" la exposición se abrirá automáticamente en una ventana del navegador.

Exposición personalizada

Esta opción requiere cargar los datos de la colección, además de seleccionar el directorio donde se desea albergar ("directorio de instalación"). Para ello, simplemente se han de seleccionar las ubicaciones donde se encuentran el CSV con los metadatos, así como la carpeta de los objetos digitales, y cubrir el resto de parámetros solicitados (título, descripción, ...).

Una vez completados los campos deseados (nótese que el único obligatorio es el correspondiente al directorio de instalación), solo queda pulsar sobre el botón "Generar" y esperar a que se complete el proceso, lo cual tomará un par de minutos. Una vez finalizado, al igual que en el caso anterior, se mostrará el mensaje "¡Exposición generada correctamente!" y, al pulsar sobre el botón "Ok", esta se abrirá automáticamente en una ventana del navegador.

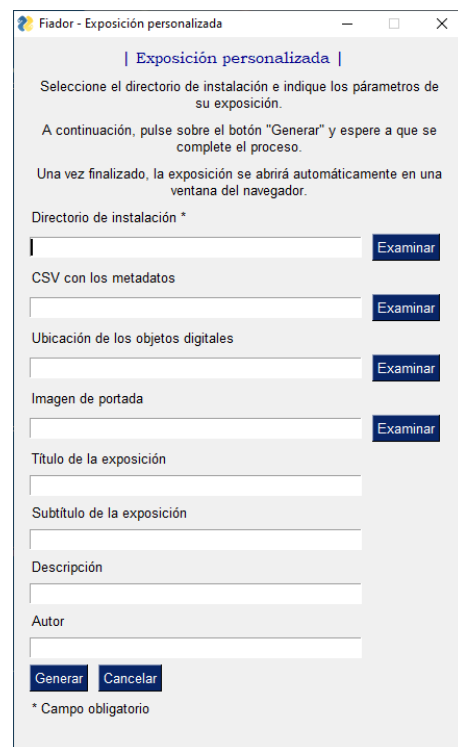
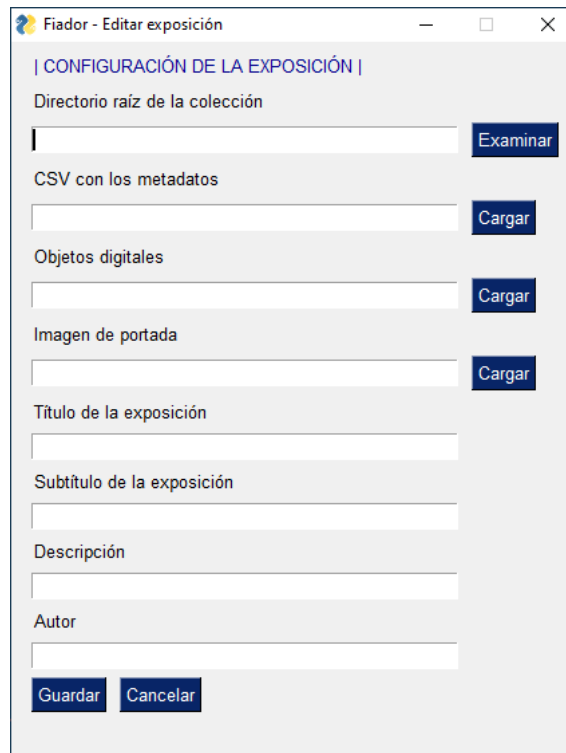


Figura 6 - Fiador 1.0. Ventana "Exposición personalizada"

4.3.2. Editar una exposición

La edición de una exposición es muy similar al proceso de generación de una exposición personalizada, pues en ambos casos se solicitan los mismos datos. Una vez cubiertos los campos deseados, simplemente se habrá de pulsar sobre el botón “Guardar” y esperar a que se complete el proceso. Al igual que en los casos anteriores, una vez finalizado se mostrará el mensaje "¡Exposición generada correctamente!" y, al pulsar sobre el botón "Ok", se abrirá automáticamente en una ventana del navegador.



The screenshot shows a web application window titled "Fiador - Editar exposición". The main heading is "| CONFIGURACIÓN DE LA EXPOSICIÓN |". The form contains the following fields and buttons:

- Directorio raíz de la colección: Input field with an "Examinar" button.
- CSV con los metadatos: Input field with a "Cargar" button.
- Objetos digitales: Input field with a "Cargar" button.
- Imagen de portada: Input field with a "Cargar" button.
- Título de la exposición: Input field.
- Subtítulo de la exposición: Input field.
- Descripción: Input field.
- Autor: Input field.

At the bottom of the form are two buttons: "Guardar" and "Cancelar".

Figura 7 - Fiador 1.0. Ventana "Editar exposición"

4.3.3. Abrir una exposición

La última funcionalidad de *Fiador* simplemente ejecuta el sitio web sobre el servidor *Jekyll* local, que una vez iniciado, lo abrirá automáticamente en una ventana del navegador.

4.3.4. Publicar una exposición

La primera vez que se genera la exposición, tanto si es de prueba como personalizada, *Jekyll* creará automáticamente una nueva carpeta denominada `_site` que contiene el sitio web completo generado a partir de la plantilla y los datos correspondientes.

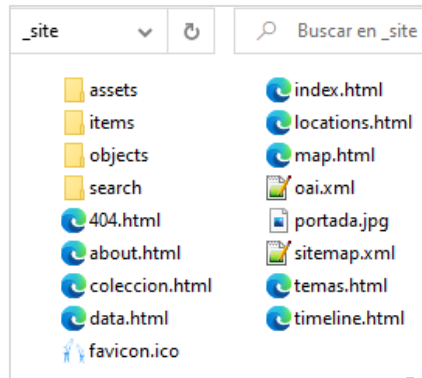


Figura 8 - Estructura de la carpeta `_site`, generada automáticamente con *Jekyll*.

Es el contenido de esta carpeta `_site` el que deberemos alojar finalmente en nuestro servidor, local o de Internet, para proceder a su publicación.

Sin embargo, antes poder publicar el contenido de la carpeta "`_site`" en el servidor elegido, es esencial regenerar las URL de los recursos para que, en lugar de apuntar a ubicaciones locales, se dirijan al servidor web de Internet. Este proceso implica la modificación de las rutas de acceso de los archivos y enlaces dentro del contenido para que se correspondan con las direcciones públicas en línea.

Por el momento, este proceso se debe llevar a cabo editando manualmente el archivo `_config.yml` de la plantilla del sitio web, para que el parámetro "url" apunte a la localización exacta del servidor web y volviendo a regenerar la colección, lo cual se puede llevar a cabo mediante la opción "Abrir colección" de *Fiador*. Una vez seguido este proceso ya se podría subir el contenido de la carpeta `_site` al servidor.

Téngase en cuenta que cada vez que se edita la colección, algunos ficheros de esta carpeta variarán, dando lugar a una versión actualizada con los cambios, por lo que se deberán volver a cargar en el servidor correspondiente, para su publicación en línea.

4.4. Prueba de concepto

7.4.1. Definición temática

Como prueba de concepto se creó un servicio de información patrimonial, bajo estructura de museo virtual, o *cibermuseo*, sin referente real, que permitiese contribuir a la catalogación, investigación y difusión de diversos documentos históricos relacionados con la comarca de Ferrol. Dicho acervo se encuentra integrado por recursos documentales de diversa procedencia, tipologías, temáticas y coberturas temporales, que se constituyen en valiosas fuentes primarias y que son desconocidas para el público general.

4.4.2. Caracterización del *dataset*

El conjunto de datos utilizados comprende 17 objetos digitales que representan documentos históricos de diversa índole, dentro del periodo comprendido entre 1806 y 1938, cada uno de cuales cuenta con su propio conjunto de atributos.

Estos 17 objetos digitales son descritos mediante un total de 23 registros y presentados en el sitio web en forma de 12 recursos. La no concordancia entre estos números se debe a que algunos de los objetos representan el anverso de una postal, por lo que se ha decidido tratarlos como "objetos compuestos". Lo mismo ocurre con varios números de una misma revista que se han decidido tratar en formado de compendio. De acuerdo al modelo de la base de datos estos "objetos compuestos" requieren de un registro concreto en el que se expresen las relaciones entre ellos y, en base al diseño de la plantilla, se presentan agrupados en un único recurso, de ahí la divergencia entre el número de objetos digitales, el número de registros de la base de datos y el número de recursos presentados en el sitio web.

Cabe señalar, además, que la calidad y cantidad de los datos no se ha tenido en cuenta a la hora de dar forma al *dataset*, en tanto la prueba de concepto se enfoca en evaluar el marco de trabajo y el modelo de publicación en sí mismo, no la abundancia o la integridad de los datos individuales. En otras palabras, el objetivo principal es validar la viabilidad del servicio de información patrimonial y su estructura como *cibermuseo*, así como si cumple con el propósito de catalogar y divulgar documentos históricos. Por lo tanto, la prueba de concepto se centra en la tecnología y la infraestructura necesarias para el servicio, y no en la perfección o exactitud de los datos.

4.4.3. Implementación

Una vez generada la exposición de forma local utilizando *Fiador*, se procedió a editar el archivo `_config.yml` de la plantilla, para que el parámetro "url", apuntase a la localización exacta del servidor web elegido. A continuación, se utilizó *Fiador* para abrir la exposición y que, de este modo, se regenerasen todos los enlaces de la carpeta `_site`.

La publicación del sitio web se llevó a cabo utilizando la plataforma de alojamiento web gratuito, *ooowebhost*⁶³, bajo el nombre de dominio <https://memoriaferrolterra.ooowebhostapp.com/>.

Asimismo, se utilizó *FileZilla*⁶⁴ como herramienta de transferencia de archivos, a fin de cargar el contenido completo de la carpeta *_site* en el servidor web de *ooowebhost*.

4.4.4. Producto resultante

Como resultado de la aplicación del marco de trabajo propuesto se ha obtenido un sitio web altamente funcional y atractivo, que destaca por ser completamente responsivo, garantizando una experiencia uniforme desde cualquier dispositivo. Además, su interfaz limpia y alineada con las tendencias de diseño del momento no solo le otorga una apariencia visualmente agradable, sino que también mejora su usabilidad. La navegación intuitiva y fluida en el sitio hace que los usuarios puedan encontrar la información que buscan de manera eficiente, lo cual redundará en una experiencia de usuario gratificante y en la posibilidad de retener a la audiencia de manera efectiva.

A continuación, se presentan una serie de capturas de pantalla del sitio web resultante que ponen de manifiesto su capacidad para adaptarse de manera efectiva a diferentes tamaños de pantalla y que además ilustran la sencillez y facilidad de uso de su interfaz.

⁶³ *ooowebhost* (<https://www.ooowebhost.com/>) es un servicio de alojamiento web que cuenta con un plan de uso gratuito, aunque limitado en características y soporte técnico. A diferencia de otros servicios de este tipo, no fuerza la inclusión de anuncios en los sitios web alojados, aunque sí coloca una marca de agua en la parte inferior de las páginas. Cuenta con un panel de control muy fácil de usar que simplifica la gestión del sitio web, no obstante, también tiene sus inconvenientes. El rendimiento de los servidores es muy irregular en comparación con la mayoría de hostings de pago, lo cual puede impactar en la velocidad y disponibilidad del sitio. En resumen, el plan gratuito de *ooowebhost* puede ser una opción adecuada para proyectos pequeños o de pruebas, pero no es recomendable para aquellos más grandes o críticos.

⁶⁴ *FileZilla* (<https://filezilla-project.org/>) es un popular cliente de código abierto para la transferencia de archivos entre un equipo local y un servidor remoto y es compatible con una variedad de protocolos de transferencia, como FTP, SFTP y FTPS.

FerrolTerrana
Creada con Fíador 1.0

Inicio Colección Temas Línea de tiempo

Buscar en la colección

Bienvid@s
Este espacio digital recoge una colección de diversos documentos históricos relacionados con la comarca de Ferrol.
[Ver colección »](#)

Buscar en la colección
Introduce los términos de búsqueda [Buscar](#)
[Operadores de búsqueda](#)

Periodo abarcado
De 1806 a 1938
[Ver línea de tiempo](#)

Temas destacados
[ferrol](#) [tarjeta postal](#) [fototipia](#)
[arsenal](#) [fene](#) [perlio](#) [valdoviño](#)
[Ver temas](#)

Tipología de la colección
2 fotografías 🖼️
2 dibujos 🖋️
1 grabado 📖
1 libro 📚
1 audio 🎧
4 postales 📮
1 recopilación 📁
[Ver Colección](#)

Metadatos de la colección
[Colección completa CSV](#) [Colección completa JSON](#) [Línea de tiempo JSON](#) [Facetas JSON](#) [Código fuente](#)

Última actualización: 13/09/2023
2023 | Lorena P.C.

Figura 9 – FerrolTerrana. Página de inicio.


FerrolTerra
Creada con Fíador 1.0

Inicio Colección Temas Línea de tiempo 🔍


Explorar la colección

[Filtrar](#) [Limpiar](#) [Orden aleatorio](#) [Búsqueda avanzada](#)


12 de 12 recursos




Ferrol: Parte del Arsenal
Fecha: 1870
Tema: [Ferrol](#) [arsenal](#)
[FOTOGRAFIA](#)
[Ver recurso](#)




Cenotafio colocado en la Iglesia de San Julián
Fecha: 1806
Autor: Santiago, Jose María de
Tema: [Ferrol](#) [Iglesia](#) [San Julián](#)
[cenotafio](#)
[GRABADO](#)
[Ver recurso](#)




A la ciencia
Fecha: 1880
Autor: Lasso de la Vega, Ángel
Tema: [ciencia](#) [poesía](#) [oda](#)
[LIBRO](#)
[Ver recurso](#)




Postal Perlió
Fecha: 1905
Autor: Casa Leira
Tema: [Ferro](#) [Perlió](#) [tarjeta postal](#)
[fotografía](#)
[POSTAL](#)
[Ver recurso](#)




Cabo Prioriño entrando al Ferrol
Fecha: 1863
Autor: Monleón y Torres, Rafael
Tema: [Ferro](#) [cabo](#) [Prioriño](#)
[DIBUJO](#)
[Ver recurso](#)




Postal Arsenal de Ferrol
Fecha: 1905
Autor: Casa Leira
Tema: [Ferro](#) [arsenal](#) [tarjeta postal](#)
[fotografía](#)
[POSTAL](#)
[Ver recurso](#)




Ferrol Ag. 81
Fecha: 1863
Autor: Monleón y Torres, Rafael
Tema: [Ferro](#)
[DIBUJO](#)
[Ver recurso](#)




Revista Navegación
Fecha: 1918-1919
Autor: Amich Bert, Julián
[RECOPILACION](#)
[Ver recurso](#)




Postal río Jubia
Fecha: 1905
Autor: Casa Leira
Tema: [Narón](#) [Jubia](#) [río Jubia](#)
[tarjeta postal](#) [fotografía](#)
[POSTAL](#)
[Ver recurso](#)



Procesión Virgen del Carmen
Fecha: 1938
Tema: [Ferro](#) [procesión](#) [patrona](#)
[fecha](#) [virgen del Carmen](#) [puerto](#)
[FOTOGRAFIA](#)
[Ver recurso](#)



Tango del Ferrol
Fecha: 1897
Autor: El Mochuelo; Rodríguez, Joaquín (el Hijo del Ciego) - int.
Tema: [tango](#)
[AUDIO](#)
[Ver recurso](#)



Postal Playa de Valdoviño
Fecha: 1905
Autor: Casa Leira
Tema: [Valdoviño](#) [Frasqueira](#) [playa](#)
[tarjeta postal](#) [fotografía](#)
[POSTAL](#)
[Ver recurso](#)

Última actualización: 13/09/2023
2023 | Lorena P.C.


Figura 10 - FerrolTerra. Página "Colección".


FerrolTerrana
Creada con Fíador 1.0

Inicio Colección Temas Línea de tiempo


Línea de tiempo


| De 1806 a 1938 | Año







1806 

1863 



1870 

1880 **A la ciencia**


1897 **Tango del Ferrol**


1905 **Postal Arsenal de Ferrol**  
Postal Perlio  
Postal Playa de Valdoviño  
Postal río Jubia  

1918 **Revista Navegación**  **Revista Navegación 25/11/1918** 

1919 **Revista Navegación**  **Revista Navegación 01/01/1919**  **Revista Navegación 15/01/1919** 

1938 

Última actualización: 13/09/2023
2023 | Lorena P.C.

Figura 11 - FerrolTerrana. Página "Línea de tiempo".



Figura 12 - FerrolTerrana. Página "Temas".

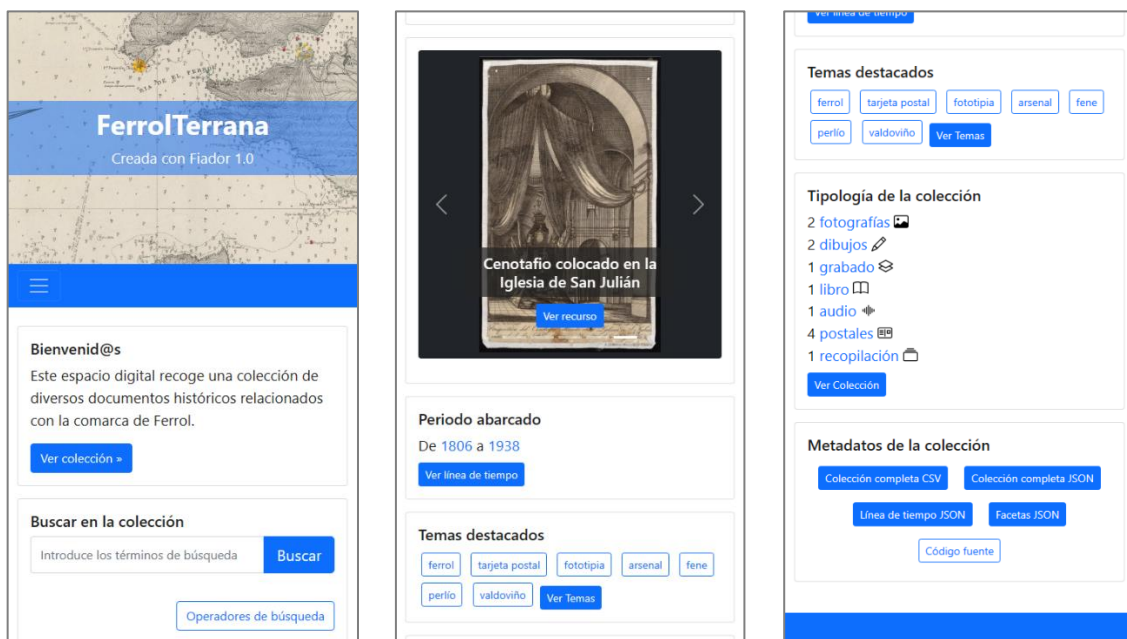


Figura 13 - FerrolTerrana. Distintas vistas de página desde dispositivo móvil.

5. Resultados y discusión

Como resultado de esta investigación, se ha propuesto un marco de trabajo que tiene como objetivo simplificar y automatizar el proceso de creación y publicación de colecciones digitales, garantizando la accesibilidad y la usabilidad para un público más amplio.

Este marco de trabajo se basa en la filosofía de la mínima computación, la cual busca utilizar los menores recursos posibles en términos de software, hardware y conocimientos informáticos. En este contexto, se identificó que *CollectionBuilder-CSV* se ajustaba bien este concepto, ya que no requiere un entorno de servidor complejo para su funcionamiento. Cumple, además, con estándares internacionales de catalogación y ofrece diferentes niveles de búsqueda para los usuarios. Sin embargo, requiere el uso de *Ruby* y la línea de comandos, lo que dificulta su uso enormemente.

Para superar estas limitaciones, se desarrolló un software llamado "Fiador", escrito en *Python* y de código abierto, que simplifica y automatiza el proceso de vinculación de objetos digitales con los metadatos en formato CSV y la posterior generación automática de los archivos necesarios para el sitio web. Su interfaz gráfica lo hace accesible incluso para usuarios sin experiencia técnica y, además, ofrece la opción de generar una exposición de prueba, con los datos incluidos en la propia plantilla, lo que sin duda facilita aún más si cabe su uso.

De este modo, el flujo de tareas para la implementación del marco de trabajo propuesto quedó configurado por tres sencillos pasos: copiar la plantilla de metadatos, reunir los objetos digitales y utilizar *Fiador* para generar el sitio web.

Como prueba de concepto, se creó un servicio de información patrimonial, sin referente real, relacionado con la comarca de Ferrol. En este contexto, el producto web resultante se posiciona como una excelente forma de divulgar este tipo de patrimonio documental, en tanto, los contenidos son susceptibles de ser fácil y periódicamente actualizados e incrementados, generando, además, un lenguaje de comunicación multimedia que, aparte de resultar atractivo visualmente, permite ofrecer información completa y de calidad. Pero no solo eso, sino que es capaz de proporcionar experiencias personales interactivas, lo cual da la oportunidad al visitante de centrarse en sus temas de interés y de establecer un diálogo con la colección.

Cabe señalar, que la magnitud de la inversión en términos de tiempo es mucho menor a la de las otras soluciones estudiadas, ya que los costes de su puesta en marcha son significativamente inferiores. Además, el hecho de ser y estar basado en software libre, permite planteamientos cada vez más ambiciosos, sugerentes y efectivos, haciendo que la propuesta de este marco de trabajo pueda ser más sólida a medida que pasa el tiempo.

Si bien *Fiador* ha demostrado ser una herramienta valiosa, no representa una solución perfecta. Una de las áreas de mejora evidentes radica en la necesidad de depurar el código para garantizar su eficacia y seguridad. Además, otro hándicap importante a considerar es que *Fiador* es exclusivamente compatible con *Windows*. Esta limitación restringe su accesibilidad y utilidad

para aquellos usuarios que utilizan otros sistemas operativos, como *MacOS* o *Linux*. Para lograr un alcance más amplio y una mayor versatilidad, sería fundamental desarrollar versiones compatibles con múltiples plataformas, lo que permitiría a un público más diverso aprovechar las ventajas de esta herramienta.

Resulta evidente también, que la necesidad de tener que editar el archivo *_config.yml* manualmente para poder llevar a cabo la publicación en el servidor final, se destaca como otra de las principales áreas de mejora que deben abordarse en futuras versiones de *Fiador*. La edición manual de este archivo puede resultar en un proceso propenso a errores y poco eficiente, ya que requiere que los usuarios realicen ajustes precisos sin la comodidad de una interfaz amigable. En este sentido, una mejora crucial sería la implementación de una interfaz de usuario más intuitiva y accesible que permita la configuración de parámetros de manera más sencilla y sin necesidad de intervención directa en el archivo de configuración. Esto no solo simplificaría el proceso, sino que también reduciría el riesgo de errores humanos y mejoraría la experiencia del usuario al utilizar *Fiador*.

En esta misma línea, el programa también podría beneficiarse de futuras actualizaciones que permitiesen la carga automática de los archivos al servidor web elegido, lo cual agilizaría considerablemente el flujo de trabajo y mejoraría la experiencia del usuario. Estas mejoras potenciales podrían hacer de *Fiador* una herramienta aún más integral y efectiva en el ámbito de la gestión de proyectos y del desarrollo web.

En resumen, si el objetivo de este marco de trabajo era simplificar y automatizar la creación y publicación de colecciones digitales, siguiendo principios de computación mínima y garantizando la accesibilidad para un público más amplio, no cabe duda de que *Fiador*, como herramienta complementaria, supera las limitaciones técnicas de *CollectionBuilder-CSV*, haciéndolo más accesible y fácil de usar para usuarios no técnicos; no obstante, representa todavía un punto de partida que necesita seguir siendo explorado y mejorado, cuestión en la que cualquier persona interesada está invitada a participar.

6. Conclusiones

El marco de trabajo propuesto ha demostrado ser una solución sencilla y efectiva para la creación y publicación de colecciones digitales. Su enfoque en la computación mínima, la accesibilidad y la usabilidad, así como su capacidad para adaptarse a diferentes contextos y tipos de colecciones, lo convierten en una herramienta a considerar en el campo de la gestión del patrimonio cultural.

La prueba de concepto, que resultó en la creación de un museo virtual sobre documentos relacionados con la comarca de Ferrolterra, ilustra la aplicabilidad y el potencial de esta solución en la catalogación, investigación y difusión del patrimonio histórico y cultural. A medida que *Fiador* y el marco de trabajo evolucionen, es posible que se conviertan en herramientas a tener en cuenta para una amplia gama de instituciones y proyectos relacionados con el patrimonio.

En resumen, este trabajo de investigación ha establecido una base sólida para la creación y gestión de colecciones digitales, abordando desafíos técnicos y de accesibilidad. La combinación de *CollectionBuilder-CSV* y *Fiador* ofrece una solución eficiente para la creación y publicación de colecciones digitales, cumpliendo con los principios de computación mínima y facilitando el acceso al patrimonio cultural en línea. Esta aproximación simplifica la vinculación de objetos digitales con metadatos y la generación del sitio web, haciéndolo asequible incluso para aquellos sin experiencia en programación. Esta herramienta tiene el potencial de democratizar la preservación y difusión del patrimonio cultural al hacer que sea más fácil y accesible para una amplia audiencia. Sin embargo, se deben abordar desafíos potenciales, como la necesidad de mantener y mejorar su código fuente o de asegurar de que sea compatible con diferentes sistemas operativos.

7. Referencias bibliográficas

- Daudinot Fournier, Isabel. "Organización y recuperación de información en Internet: teoría de los metadatos." *ACIMED* 14, n.º. 5 (2006). http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S1024-94352006000500006&lng=es&nrm=iso&tlng=es
- Dempsey, Lorcan. "Scientific, Industrial, and Cultural Heritage: A Shared Approach." *Ariadne*, no. 22 (1999). <http://www.ariadne.ac.uk/issue/22/dempsey/>
- Directiva (UE) 2019/790 del Parlamento Europeo y del Consejo, de 17 de abril de 2019, sobre los derechos de autor y derechos afines en el mercado único digital y por la que se modifican las Directivas 96/9/CE y 2001/29/CE. (DOUE núm. 130, de 17 de mayo de 2019, pp. 92-125). <https://www.boe.es/buscar/doc.php?id=DOUE-L-2019-80826>
- Directiva 2001/29/CE del Parlamento Europeo y del Consejo, de 22 de mayo de 2001, relativa a la armonización de determinados aspectos de los derechos de autor y derechos afines a los derechos de autor en la sociedad de la información (DOCE núm. 167, de 22 de junio de 2001, pp. 10-19). <https://www.boe.es/buscar/doc.php?id=DOUE-L-2001-81549>
- García Marco, Francisco Javier. "De la catalogación de patrimonio a los esquemas de metadatos." *Boletín de la ANABAD* 64, n.º 4 (2014): 21-77. <https://www.anabad.org/wp-content/uploads/2014/11/2014.4.pdf>
- García Melero, Luis Ángel. "La biblioteca digital revisitada." *Boletín de la ANABAD* (2009). http://eprints.rclis.org/13628/1/LA_BIBLIOTECA_DIGITAL_REVISITADA.pdf
- ISO 12234-2:2001. "Electronic still-picture imaging - Removable memory - Part 2: TIFF/EP image data format."
- ISO/IEC 10918-1:1994. "Information technology - Digital compression and coding of continuous-tone still images: Requirements and guidelines."
- Kahn, Robert, y Robert Wilensky. "A Framework for Distributed Digital Object Services" (1995). <http://www.cnri.reston.va.us/k-w.html>
- Moro Cabero, Manuela. "Identificación, caracterización y selección de formatos para la preservación del recurso digital." *Métodos de Información* 9, n.º 16 (2018): 49-90. <http://dx.doi.org/10.5557/IIMEI9-N16-049090>
- NISO. *A Framework of Guidance for Building Good Digital Collections*. 3.^a ed. Baltimore: NISO, 2007. <https://www.niso.org/sites/default/files/2017-08/framework3.pdf>
- Pérez-Hernández Dayni y Delgado-Dapena, Martha Dunia. "Modelo de gestión de objetos digitales para la gestión de soluciones tecnológicas." *Ingeniería Industrial* 34, n.º 1 (2013): 40-49. <http://www.scielo.sld.cu/pdf/rii/v34n1/rii05113.pdf>

- Rodríguez Reséndiz, Perla Olivia, y María Teresa Fernández Bajón. "Conectando los saberes de bibliotecas, archivos y museos (BAM) en torno a la preservación de documentos analógicos y de origen digital." México: Universidad Nacional Autónoma De México, 2019. https://ru.iibi.unam.mx/jspui/handle/IIBI_UNAM/L212
- Salse, Marina, Jornet, Núria y Javier Guallar. "El patrimonio universitario desde una perspectiva GLAM. Análisis de los sitios web de las universidades europeas". *Revista General de Información y Documentación* 31, n.º2 (2019): 521-543. <https://revistas.ucm.es/index.php/RGID/article/view/77215/4564456559308>
- Serrano Fernández, María. "La digitalización del patrimonio cultural La Directiva sobre derechos de autor y derechos afines en el mercado único digital." *Indret* 3 (2020). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3963849
- Torres Márquez, Martín. "Propuesta de 'UCOMuseo virtual', museo y metamuseo de la Universidad de Córdoba (España)." *Revista de Formación e Innovación Educativa Universitaria* 5, n.º 3 (2012): 133-152. https://refiedu.webs.uvigo.es/Refiedu/Vol5_3/REFIEDU_5_3_1.pdf
- UNE-ISO 19005-1:2008. Gestión de documentos. Formato de fichero de documento electrónico para la conservación a largo plazo. Parte 1: Uso del PDF 1.4 (PDF/A-1).
- UNESCO. "Directrices para la preservación del patrimonio digital" (2006). https://unesdoc.unesco.org/ark:/48223/pf0000130071_spa
- Wikle, Olivia. "What is static web and what's it doing in the digital humanities classroom?" *Dh Lib* (blog), 22 de junio de 2020. <https://dhandlib.org/2020/06/22/what-is-static-web-and-whats-it-doing-in-the-digital-humanities-classroom/>