



SIMURG, PATRIMONIO DIGITALIZADO DEL CSIC PARA LA HISTORIA DE LA CIENCIA

SIMURG, CSIC DIGITIZED HERITAGE FOR THE HISTORY OF SCIENCE

Autoras:

Laura Donadeo Navalón. Consejo Superior de Investigaciones Científicas.

laura.donadeo@bib.csic.es

Alexandra López Pérez. Consejo Superior de Investigaciones Científicas.

sandra@bib.csic.es

Resumen:

Se presenta Simurg, la biblioteca de fondos patrimoniales digitalizados del Consejo Superior de Investigaciones Científicas.

Desde 2008 la Red de Bibliotecas y Archivos del CSIC lleva a cabo una política de digitalización de fondos bibliográficos y archivísticos con criterios técnicos aceptados internacionalmente con el fin de que puedan ser consultados por los usuarios potenciales en cualquier parte del mundo y que los originales se conserven en las mejores condiciones posibles, reduciendo su manipulación.

A día de hoy se han gestionado más de 150 proyectos de una gran variedad documental (manuscritos, impresos antiguos, fotografías, partituras, mapas, etc.) La colección asciende a más de 25.000 documentos y 1.780.000 imágenes.

Se repasan los proyectos más representativos, los criterios de selección de documentos, la planificación, ejecución, difusión, cuestiones técnicas de interés y los retos asociados al trabajo en red.

Abstract:

Simurg, the library of digitized heritage collections of the Spanish National Research Council, is presented.

Since 2008, CSIC's Libraries and Archives Network has carried out a policy of digitizing bibliographic and archival material with internationally accepted technical criteria so that they can be consulted by potential users anywhere in the world and that the originals are preserved in the best possible conditions, reducing handling.



As of today, more than 150 projects of a great documentary variety (manuscripts, old prints, photographs, scores, maps, etc.) have been managed. The collection grew to more than 25,000 documents and 1,780,000 images.

The most representative projects, document selection criteria, planning, execution, dissemination, technical issues of interest and the challenges associated with networking are reviewed.

Palabras clave:

Bibliotecas digitales, Simurg, Red de Bibliotecas y Archivos del CSIC

Keywords:

Digital libraries, Simurg, CSIC's Libraries and Archives Network

INTRODUCCIÓN

Simurg es la colección de fondos patrimoniales del CSIC digitalizados según el *Plan Director de Digitalización del CSIC*. El nombre proviene de la mitología persa, donde *Simurg* es el nombre del pájaro inmortal que anida en las ramas del árbol de la Ciencia. Desde 2008, las actuaciones se han enmarcado en los distintos planes estratégicos del Organismo y a día de hoy, se han gestionado más de 150 proyectos de una gran variedad documental (manuscritos, impresos antiguos, fotografías, partituras, mapas, etc.) La colección asciende a más de 25.000 documentos y 1.780.000 imágenes, que se pueden consultar en el portal [Simurg, fondos digitalizados del CSIC](#).

PRINCIPALES HITOS

Antes de la puesta en marcha de la política de digitalización común para la Red de Bibliotecas, hubo distintas iniciativas en la Institución que han permitido ofrecer imágenes y textos de interés. Estas digitalizaciones tuvieron como objetivo principal la difusión de los fondos, obteniendo una buena aceptación entre los usuarios, pero no consideraron los aspectos relativos a la preservación de las propias imágenes digitales.

En 2008 se incorporó una persona dedicada a poner en marcha y desarrollar los temas de digitalización. En 2009, se elaboró el [Manual Técnico de Digitalización](#) y en 2010 el [Plan Director para la digitalización de fondos del CSIC](#). La aplicación de esta teoría se transformó en una gestión uniforme y estandarizada de los datos, de la estructura de ficheros, de los nombramientos de imágenes, de los ficheros técnicos de gestión y preservación, etc.

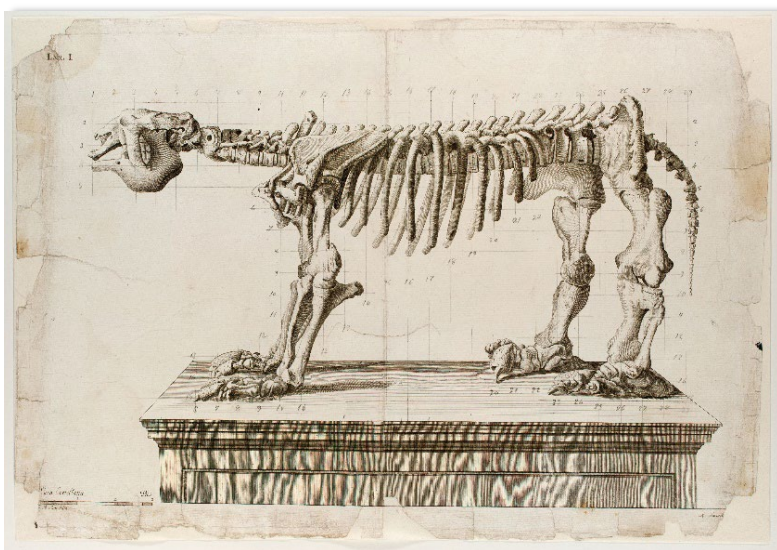


Figura 1.
Esqueleto del megaterio montado (con cola) - Megatherium americanum. Archivo del Museo Nacional de Ciencias Naturales. Signatura: ACN110B/004/05641

En 2010 se adquirió [Goobi](#), herramienta para la gestión de los proyectos de digitalización, y se configuró un visor específico para la difusión. Cuando se escogió, había aún menos herramientas de las que actualmente ofrece el mercado, que siguen siendo pocas. El software en sí tenía, como todos, aciertos y desaciertos. Entre los primeros, sigue siendo uno de ellos la generación de METS de forma autónoma y distribuida. Los desaciertos, lamentablemente, fueron algunos más y lastraron el proyecto, restándole frescura y dinamismo. A partir de 2011 se pudo disponer de un servidor de conservación para las imágenes que se obtienen en los proyectos y en 2013 se adquirió un escáner dedicado a Simurg, que ha permitido la continuidad de los proyectos con una menor dependencia de las partidas presupuestarias anuales.

Desde 2015 Simurg es recolectado por Hispana, agregador español de contenidos digitales y desde 2017 por Europeana, agregador europeo de proyectos de digitalización de patrimonio.

En 2020 se adquiere el software [Limb Processing](#) para la gestión y el procesamiento de imágenes y metadatos, así como la herramienta Open Access LIBSAFE Discovery, de Libnova, que permitirá en breve la implantación de una nueva interfaz de consulta de los fondos patrimoniales digitalizados así como un nuevo sistema de gestión de los objetos digitales.

COLECCIONES DESTACADAS

Todos los proyectos de digitalización llevados a cabo son importantes por el contenido o por las características de los documentos; se pueden destacar algunos por su relevancia especial:



Obras antiguas Química y Alquimia.

Primer proyecto de Simurg, con más de 400 obras custodiadas en distintas bibliotecas de la Red; se diseñó una web específica para su difusión.



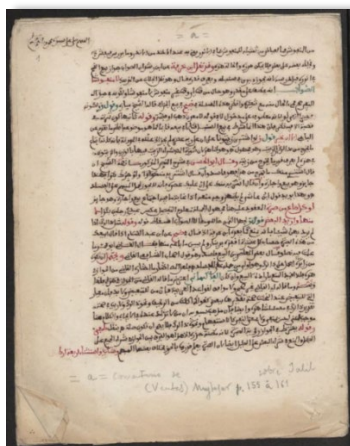
Manuscript@.

Reúne los 153 manuscritos en hebreo, árabe, aljamiado, persa y turco conservados en las bibliotecas del CSIC. Se pueden consultar en una web dedicada, que incluye las descripciones codicológicas, bibliografía temática y enlaces de interés.

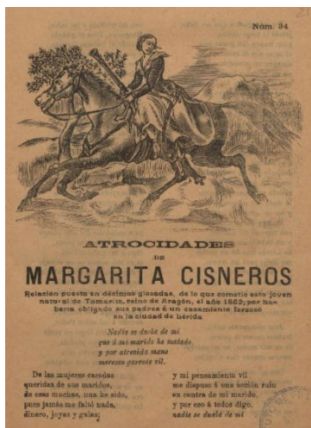


Catálogo Monumental de España.

40 piezas que suponen cerca de 40.000 imágenes de unos documentos, en parte inéditos, que se custodian en la Biblioteca Tomás Navarro Tomás y que son muy valiosos para el conocimiento del patrimonio artístico español.



Fondo Antiguo, de las bibliotecas del Museo Nacional de Ciencias Naturales, de la Estación Experimental de Aula Dei, de la Escuela de Estudios Árabes y de la Institución Milá y Fontanals, con un total de 699 obras que forman un importante conjunto de ediciones históricas relacionadas con distintos ámbitos de las Ciencias y las Humanidades.



Plegos de cordel y Libros de viajes con, respectivamente, 312 documentos de gran interés para el estudio de la literatura popular y 152 obras en distintos idiomas; ambos conjuntos pertenecen a la Biblioteca Tomás Navarro Tomás.

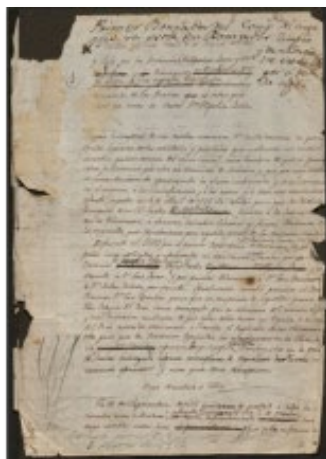


Archivo Etnográfico y Folklórico de Catalunya Reúne 1.797 fotografías digitalizadas que se conservan en la Institución Milá y Fontanals con un enorme valor antropológico e histórico.



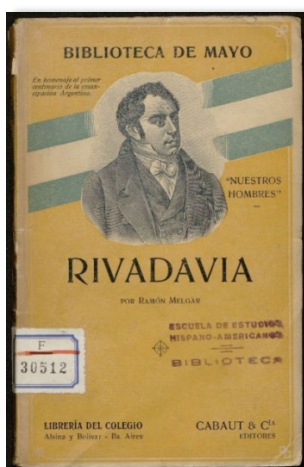
Calcos y dibujos rupestres.

Cerca de 2.000 piezas del archivo del Museo Nacional de Ciencias Naturales que son fundamentales para estudios arqueológicos y paleontológicos de la Península Ibérica. En este proyecto se ha colaborado con el Instituto del Patrimonio Cultural de España para la restauración.



Comisión Científica del Pacífico

La Comisión fue la principal empresa científica ultramarina de la España isabelina, gracias a la cual se formaron importantes colecciones zoológicas, botánicas, geológicas, antropológicas y arqueológicas, así como numerosa documentación ahora digitalizada.



Historia de América y obras con papel acidificado de la Biblioteca Americanista de Sevilla.

La biblioteca alberga una imponente colección de obras editadas en América y sobre la región, además de un basto fondo bibliográfico que se está digitalizando de forma prioritaria por su riesgo de desaparición por la acidez del papel.

LA GESTIÓN DE PROYECTOS DE DIGITALIZACIÓN

Los fondos bibliográficos o archivísticos, textuales o iconográficos, que se quieran digitalizar han de estar siempre en dominio público. Para su selección se valora que tengan un gran uso; que sean de especial relevancia; que tengan un elevado riesgo de deterioro; que complementen y refuercen líneas de investigación en desarrollo dentro del CSIC; que sean unidades temáticas que formen un conjunto de presentación, exposición, etc.; que respondan a una efeméride o acontecimiento en la sociedad o que participen en proyectos o iniciativas de digitalización de otras instituciones, tanto nacionales como internacionales.

Tradicionalmente existen dos vías para la publicación de contenidos en Simurg. Por una parte encontramos las convocatorias anuales por parte de la Unidad e Recursos de Información Científica para la Investigación (URICI-CSIC), responsable de la digitalización del patrimonio institucional. Por otra parte, atendemos desideratas de investigadores, grupos de investigación o ciudadanos.



Los requisitos para acometer los trabajos de digitalización incluyen una justificación de la necesidad o importancia de digitalizar el fondo, que los documentos estén correctamente catalogados, la comprobación de que no se encuentren ya disponibles en otras bibliotecas digitales de libre acceso y una descripción lo más ajustada posible de las características físicas (número de páginas, dimensiones, anexos, despleables, estado de conservación...)

Conviene destacar la complejidad que comporta esta gestión al trabajar en una red con 58 bibliotecas y 16 archivos. Trabajar en red nos dota de una riqueza inmensa, pero le añade complejidad a cualquier iniciativa. Cada centro tiene su propia dirección, directrices y estrategias, y hemos de conjugar los intereses de todos. Además, trabajar de forma distribuida hace que muchos de los archivos y bibliotecas sean unipersonales, de manera que los bibliotecarios a veces no tienen tiempo ni medios para dedicarle a una labor como la digitalización y difusión de fondos que puede verse, en ocasiones, como accesoria.

Una vez que se han seleccionado los proyectos, se desarrollan los aspectos técnicos implicados y necesarios para el proceso de digitalización que ha de abarcar desde la recepción física de las obras, hasta la devolución de las mismas tras la generación de imágenes digitales de calidad. Es fundamental la determinación de los tiempos, el formato físico y estados de conservación de la obra con la que se va a trabajar. Además, un correcto nombrado del registro de obra, así como de las imágenes que lo componen, siguiendo las normas propuestas por el *Plan*, es clave para un procesamiento ágil y adecuado.

Los criterios técnicos de digitalización siguen las normas pautadas en el *Plan* ya comentado en relación al formato principal de salida (TIFF en color de 400ppp) y metodología de escaneado que define las características de esta: una imagen por página y márgenes que permitan apreciar el material como objeto (cubiertas, estado de conservación, etc.).

Sin embargo, han de primar unos preceptos generales a la hora de abordar una correcta manipulación de la documentación patrimonial que garanticen en todo momento la integridad de las obras. Para ello una adaptación y ajuste adecuado de los materiales a las medidas del escáner utilizado y la búsqueda de una imagen óptima que de fe del original (páginas estiradas, encuadre, enderezamiento, nitidez, etc.) han de alzarse como puntos esenciales en toda digitalización. El estado de conservación marca el especial



cuidado en la manipulación, ya que aquellas obras con procesos graves de humedad o el uso de papel ácido y/o extremadamente delicado, por ejemplo, requieren una mayor atención por parte del técnico.

Llegados a este punto, en el plan de trabajo convergen tanto los proyectos realizados en URICI como los que llegan de empresas externas. A partir de aquí se lleva a cabo un control de calidad sobre las imágenes resultantes, se crean los mapas estructurales sobre las obras, que ayudan en una navegación rápida por el documento; se generan los ficheros METS y los formatos derivados de los objetos digitales: PDF con OCR para consultas, JPG para publicación en Redes Sociales y web, etc.

La última fase conlleva en primer lugar, el depósito de los ficheros máster en el servidor de conservación. A continuación se procede a la publicación en el portal Simurg, que actúa como puerta de entrada a los objetos digitales, y a la publicación de los enlaces correspondientes en el catálogo colectivo del CSIC.

NUEVA HERRAMIENTA DE GESTIÓN, NUEVAS FUNCIONALIDADES

Tal como decíamos al principio, en los últimos años estamos haciendo un esfuerzo por implementar mejoras tanto en nuestra gestión interna de los proyectos, como en la forma de difundirlos y preservarlos.

Para ello, el primer paso fue comenzar a utilizar un software específico de procesado de imágenes y metadatos. Con Limb Proccesing, tareas como la generación del máster, enderezados y recortados, retoques de imagen, generación de PDF y OCR o incluso la generación de metadatos, son ahora procesos altamente automatizables.

En segundo lugar, desde 2021 trabajamos en la migración de nuestros contenidos a una nueva herramienta para la difusión, que permita además nuevas funcionalidades: búsqueda en el texto completo de los documentos, variedad de formatos de presentación, visor de documentos con IIIF. Seguimos el camino de otras instituciones que ya han puesto en marcha con éxito esta herramienta, como la [Biblioteca del Real Jardín Botánico \(CSIC\)](#) o la [Biblioteca Mazarina](#).

Durante este año y el siguiente pretendemos también mejorar la preservación a largo plazo de las colecciones de Simurg. Desde 2019 se viene trabajando en fijar una política de preservación de colecciones digitales documentales que incluya tanto las colecciones del repositorio institucional [Digital.CSIC](#), como las colecciones de fondos patrimoniales digitalizados en Simurg. Se va a implantar un software específico de preservación digital,

acorde con el modelo OAIS, con la idea de gestionar riesgos, desde los más básicos como los fallos en los soportes de almacenamiento, hasta los más complejos, como los formatos que quedan obsoletos o la documentación de los contenidos.

EL OBJETIVO FINAL: DIFUSIÓN Y REUTILIZACIÓN

La principal razón de ser de todos estos trabajos de digitalización es sin duda garantizar el acceso y difusión de los contenidos, no solo a investigadores de cualquier materia, especialmente los dedicados a la Historia de la Ciencia, sino también a toda la sociedad.



Figura 2.
Cubierta del calendario Simurg 2022

De una parte, apostamos por un uso de las redes sociales que acerque la ciencia a “los no entendidos” porque, muchas veces, la belleza de la documentación del portal Simurg simplemente se disfruta, aunque se difunda con un fin didáctico. Así, celebramos efemérides, días internacionales o eventos de actualidad. Además, elaboramos productos sobre temas específicos, los llamados [Simurg Temáticos](#), donde

agrupamos documentación dispersa de diferentes bibliotecas o archivos bajo un hilo conductor. Por último, determinados proyectos merecen, por su envergadura, disponer de un portal dedicado donde se enriquecen con datos bibliográficos, reseñas, etc. Es el caso del Catálogo monumental de España, la Comisión Científica del Pacífico o Manuscript@, que ya se han mencionado anteriormente.

De otra parte, incentivamos proyectos de reutilización de nuestros contenidos, especialmente en el ámbito de las Humanidades Digitales. Así, por ejemplo, alentamos la participación ciudadana a través del proyecto de descripción colaborativa [Apadrina un documento](#), de la Biblioteca Tomás Navarro Tomás, o trabajamos en la edición digital de los contenidos de Simurg en formatos abiertos como el epub. Destaca el portal [Mapping Pliegos](#), que ofrece una base de datos y biblioteca digital de los pliegos sueltos publicados en España entre los siglos XIX y XX. El proyecto es fruto de la colaboración entre investigadores del campo de la literatura popular impresa, bibliotecarios y documentalistas, e instituciones. En 2022, además, estamos colaborando con el grupo



Música y Sociedad (IMF-CSIC) para digitalizar parte del fondo con el objetivo de explorar procedimientos de OMR (Optical Music Recognition) para capturar las melodías digitalizadas y codificarlas en MusicXML para su posterior renderización en notación moderna e incorporación al [fondo de Música Tradicional](#).

Bibliografía:

- COTTEREAU, M., 2016. Simurg, Fondos Digitalizados del CSIC, resumen de actividades entre los años 2008 y 2016. Enredadera [en línea], no. 28, pp. 1–3. [Consulta: 6 April 2022]. DOI 10.20350/digitalCSIC/9829. Disponible en: <https://digital.csic.es/handle/10261/140848>.
- DONADEO NAVALÓN, L. and LÓPEZ PÉREZ, A., 2022. Proyectos destacados en “Simurg: Fondos digitalizados del CSIC” durante el año 2021. Enredadera [en línea], no. 37, pp. 67–71. [Consulta: 6 April 2022]. DOI 10.20350/digitalCSIC/14538. Disponible en: <http://hdl.handle.net/10261/263275>.
- LÓPEZ PÉREZ, A. and COTTEREAU, M., 2021. En cualquier momento, en cualquier lugar. La digitalización del patrimonio documental como instrumento del desarrollo y la igualdad (ODS y Agenda 2030). Enredadera [en línea], no. 36, pp. 11–17. [Consulta: 6 April 2022]. DOI 10.20350/digitalCSIC/13938. Disponible en: <http://hdl.handle.net/10261/245970>.
- LÓPEZ PÉREZ, A. and DONADEO NAVALÓN, L., 2021. Los archivos como testimonios de las expediciones científicas a América [en línea]. S.l.: s.n. [Consulta: 6 April 2022]. Disponible en: <http://hdl.handle.net/10261/242313>.
- LÓPEZ PÉREZ, A. and JIMÉNEZ MARTÍN, M., 2018. 10 años de proyectos de digitalización en la Red de Bibliotecas y Archivos del CSIC: 2008-2018 [en línea]. S.l.: s.n. [Consulta: 6 April 2022]. Disponible en: <https://digital.csic.es/handle/10261/160467>.
- PONSATI OBIOLS, A., ORTEGA-MONASTERIO, M.T. and SANTAMARINA, C., 2012. Manuscripta@CSIC un proyecto piloto SIMURG: fondos digitalizados del CSIC. Novática [en línea], no. 216, pp. 27–33. [Consulta: 6 April 2022]. Disponible en: <http://hdl.handle.net/10261/51948>.
- SANTAMARINA, C., 2015. De Goobi a SIMURG: un ciclo. Enredadera [en línea], no. 25, pp. 21–23. [Consulta: 6 April 2022]. DOI 10.20350/digitalCSIC/9804. Disponible en: <https://digital.csic.es/handle/10261/116470>.