

Nueva herramienta de gestión y difusión de Simurg, fondos digitalizados de las bibliotecas y archivos del CSIC

LAURA DONADEO NAVALÓN; ALEXANDRA LÓPEZ PÉREZ.

*Unidad de Recursos de Información Científica para la Investigación,
Consejo Superior de Investigaciones Científicas (URICI-CSIC)*

Dirección de correspondencia: laura.donadeo@bib.csic.es

Fecha: 13/09/2022

RESUMEN:

Se presenta *Simurg*, la biblioteca de fondos patrimoniales digitalizados del Consejo Superior de Investigaciones Científicas, así como el proceso de migración de sistema de gestión y difusión producido en 2022.

Se repasan los hitos más significativos de su historia, los proyectos más representativos y la planificación y ejecución de proyectos de digitalización, haciendo especial hincapié en los cambios derivados de la reciente migración de herramienta de gestión y difusión.

PALABRAS CLAVE: Bibliotecas digitales, Historia de la ciencia, Simurg, Red de Bibliotecas y Archivos del CSIC

ABSTRACT:

Simurg, the library of digitized heritage collections of the Spanish National

Research Council, is presented, as well as the migration process of the management and dissemination system in 2022.

The most significant milestones in its history, the most representative projects and the planning and execution of digitization projects are reviewed, with special emphasis on the changes derived from the recent migration in management and dissemination tools.

KEYWORDS: Digital libraries, History of science, Simurg, CSIC's Libraries and Archives Network

INTRODUCCIÓN

Simurg es la colección de fondos patrimoniales del Consejo Superior de Investigaciones Científicas (CSIC) digitalizados según el *Plan Director para la Digitalización de fondos del CSIC* (Ponsati, 2010). 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 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.

Las actuaciones se han ido enmarcando 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.), ascendiendo la colección a más de 25.000 documentos y 1.780.000 imágenes, que se pueden consultar en el portal *Simurg*.

PRINCIPALES HITOS

Antes de la puesta en marcha de la política de digitalización común para la Red de Archivos y Bibliotecas, hubo distintas iniciativas en la Institución que permitieron ofrecer imágenes y textos de interés. Su objetivo principal fue la difusión de los fondos, obteniendo una buena aceptación entre los usuarios, pero sin considerar los aspectos relativos a la preservación de las propias imágenes digitales generadas.

En 2008 se incorporó una persona dedicada a poner en marcha y desarrollar los temas de digitalización; y en 2009 se elaboró el *Manual Técnico de Digitalización de la Red de Bibliotecas y Archivos del CSIC* (Santamarina, 2009). Para su elaboración se trabajó en un *corpus* teórico de varios documentos, ayudas, manuales, etc., que guiaban a las bibliotecas y archivos de la Red a la hora de emprender sus digitalizaciones, y que finalmente dio como resultado el *Plan Director para la digitalización de fondos del CSIC*. La aplicación de toda esta teoría se transformó en una gestión uniforme y estandarizada de los datos, de la estructura de ficheros, del nombrado de imágenes, de los ficheros técnicos de gestión y preservación, etc.

Los primeros proyectos se organizaron a través de códigos numéricos (P0001, P0002, etc.) y se estructuraron en carpetas que diferenciaban los archivos máster, los derivados y los metadatos; y se almacenaron en un servidor de gran capacidad con copias de seguridad incrementales y en cinta, facilitando así cualquier migración o implementación futura de un sistema de preservación.

En 2010 se adquirió *Goobi* (Intranda, s.f.), herramienta para la gestión de proyectos de digitalización, y se configuró un visor específico para la difusión. Cuando se escogió, existían aún menos herramientas de las que actualmente ofrece el mercado, que siguen siendo pocas. El software en sí tenía sus aciertos y desaciertos; entre los primeros, la generación de METS de forma autónoma y distribuida. Sin embargo, los desaciertos, lamentablemente, fueron más y lastraron el proyecto, restándole frescura. La implementación de una web estática, sin desarrollo externo ni interno, reforzó una imagen poco dinámica y apropiada a la naturaleza digital de los contenidos. Pese a sus limitaciones, la existencia del portal permitió hacer importantes proyectos y empezar a difundir la relevante documentación patrimonial custodiada en el CSIC.

A partir de 2011 se pudo disponer de un servidor de conservación para las imágenes que se obtienen en los proyectos. Este servidor, denominado CABAS, se encuentra albergado actualmente en el Instituto de Física de Cantabria, centro mixto del CSIC y la Universidad de Cantabria.

En 2013 se adquirió un escáner dedicado a *Simurg*, que permitió la continuidad de los proyectos con una menor dependencia de las partidas presupuestarias anuales. Y desde 2016 se cuenta con un técnico especializado

en imagen, de forma que se abordan nuevos proyectos de digitalización y otras mejoras.

En 2017 *Europeana*, agregador europeo de proyectos de digitalización de patrimonio, comienza a recolectar el contenido de *Simurg*, ya recolectado por *Hispana* desde 2015.

Por último, es en 2020 cuando se adquiere el software *Limb Processing* (I2S, s.f.) para la gestión y el procesamiento de imágenes y metadatos, así como la herramienta *Open Access LIBSAFE Discovery*, de Libnova, que ha permitido, ya en 2022, la implantación de una nueva interfaz de consulta de los objetos digitalizados.

COLECCIONES DESTACADAS

Todos los proyectos de digitalización llevados a cabo han sido importantes por el contenido, por las características de los documentos o por ambas cosas; aunque se pueden destacar algunos por su especial relevancia:

Obras antiguas de Química y Alquimia (Proyecto de Digitalización de Obras Antiguas de Química y Alquimia, 2011). 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@ (Manuscript@CSIC, 2010). Reúne 153 manuscritos en hebreo, árabe, aljamiado, persa y turco conservados en las bibliotecas del CSIC. Además de en *Simurg*, se pueden consultar en una web propia que incluye las descripciones codicológicas, bibliografía temática y enlaces de interés.

Catálogo Monumental de España (Catálogo Monumental de España (1900-1961), 2008). 40 piezas que suponen cerca de 40.000 imágenes de documentos, en parte inéditos, que se custodian en la Biblioteca Tomás Navarro Tomás del Centro de Ciencias Humanas y Sociales y son muy valiosos para el conocimiento del patrimonio artístico español. Se acompaña de un portal para su consulta, con una amplísima información relativa al proyecto y a la importancia de las obras.

Archivo Etnográfico y Folklórico de Catalunya (Institución Milá i

Fontanals de Investigación en Humanidades, 2015). Reúne 1.797 fotografías digitalizadas que se conservan en la Institución Milá i Fontanals, con un enorme valor antropológico e histórico.

Calcos y dibujos rupestres. (Morón, 2016). Cerca de 2.000 piezas del Archivo del Museo Nacional de Ciencias Naturales fundamentales para estudios arqueológicos y paleontológicos de la Península Ibérica. En este proyecto se colaboró con el Instituto del Patrimonio Cultural de España para la restauración y digitalización de una parte del fondo que se encontraba en un estado delicado.

Comisión Científica del Pacífico. La Comisión Científica del Pacífico 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.

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 i 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.

Teatro del siglo XIX. La Biblioteca Tomás Navarro Tomás del Centro de Ciencias Humanas y Sociales dispone de una amplia colección de textos teatrales desde el siglo XVII hasta la actualidad. De entre ellos destaca, escrita en castellano y catalán, una muestra representativa de lo que fue la actividad teatral en España entre 1831 y 1936.

LA GESTIÓN DE PROYECTOS DE DIGITALIZACIÓN

El CSIC posee un conjunto documental susceptible de ser digitalizado que está compuesto por fondos bibliográficos en dominio público y documentación archivística, tanto textual como iconográfica. Estos materiales son de gran importancia para distintas áreas del conocimiento, y muy especialmente para el estudio de la Historia de la Ciencia en España.

Para la selección de estos fondos bibliográficos y archivísticos se valora

que tengan un gran uso y un elevado riesgo de deterioro; que sean de especial relevancia; 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. Los requisitos de selección incluyen justificar la necesidad o importancia de digitalizar el fondo. Esto supone estudiar además si los materiales están bien descritos en el catálogo, comprobar que no se encuentren presentes en otras bibliotecas digitales de libre acceso y una descripción lo más ajustada posible de las características físicas. Este análisis determinará si se puede digitalizar con equipos propios o si es preferible externalizar los trabajos.

Tradicionalmente existen dos vías para el ingreso de contenidos en *Simurg*. Por una parte, se encuentran las convocatorias anuales realizadas por la Unidad de Recursos de Información Científica para la Investigación (URICI-CSIC), responsable de la digitalización del patrimonio institucional. Por otra parte, se atienden las desideratas de investigadores, grupos de investigación o ciudadanos. Desde la URICI se llevan a cabo proyectos con medios propios en nuestras instalaciones y además se acometen proyectos en las sedes de las bibliotecas a través de contratos con empresas. Ambos tipos acaban confluyendo en el momento del procesado y la publicación, que se hace de forma centralizada en la URICI.

Conviene destacar la dificultad que comporta esta gestión al trabajar en una red con casi 60 bibliotecas y 16 archivos. Trabajar en red dota de una riqueza inmensa, pero le añade complejidad a cualquier iniciativa. Cada centro tiene su propia dirección, sus directrices y estrategias, y se han de conjugar los intereses de todos. Igualmente, muchos de los archivos y bibliotecas del CSIC son gestionados de forma unipersonal con un solo técnico a su cargo, que a veces no tiene tiempo ni medios para dedicar a una labor como la digitalización y la difusión de fondos que puede verse, en ocasiones, como accesoria.

A la hora de abordar cualquier proyecto de digitalización la planificación previa es esencial. 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.

En segundo lugar, es fundamental la determinación de los tiempos, el formato físico y estado de conservación de la obra con la que se va a trabajar. Un correcto nombrado del registro de la obra, así como de las imágenes que lo componen, siguiendo las normas propuestas por el *Plan de Digitalización*, es clave para un procesamiento ágil y adecuado.

El proceso de digitalización en las instalaciones de URICI se realiza con dos equipos de digitalización: los escáneres Copy Book A2 (adquirido en 2013) y Copibook OS A2 (adquirido en 2022).

Los criterios técnicos de digitalización siguen las normas pautadas en el *Plan* ya comentado en relación al formato principal de salida (TIFF, en color y 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, de manera que se adapten a las normas ya propuestas; la creación de mapas estructurales sobre las obras que implica la realización de índices, por decirlo así, que ayuden en una navegación rápida por el documento; la creación de los ficheros METS y la generación de nuevos formatos necesarios para los diversos usos 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 y a continuación, la publicación en el portal *Simurg*, y de sus enlaces correspondientes en la *Biblioteca Virtual del CSIC*.

NUEVA HERRAMIENTA, NUEVAS FUNCIONALIDADES

En los últimos años se está haciendo un esfuerzo importante por establecer mejoras tanto en la gestión interna de los proyectos, como en la forma de difundirlos y preservarlos.

Y así desde 2020 se comenzó a trabajar en la búsqueda de nuevas herramientas tanto para la gestión como para la difusión de los fondos digitalizados. En este sentido, en ese mismo año se adquirió el software *Limb Processing* para el procesamiento de imágenes y metadatos, así como la herramienta *Open Access LIBSAFE Discovery*, como herramienta de difusión y consulta de los objetos digitales.

Los trabajos de migración del sistema comenzaron en abril de 2021 y culminaron con la puesta en producción del nuevo portal en junio de 2022. El equipo involucrado estuvo formado tanto por personal bibliotecario, como de gestión e informático de la URICI, en colaboración con la empresa suministradora del nuevo sistema.

La planificación de la migración incluyó, por una parte, el rediseño de la interfaz y la arquitectura del portal, con la elección de los elementos que se priorizaba destacar en la página de inicio, los distintos bloques, elementos gráficos, configuración del buscador, funcionalidades, etc.

En paralelo al rediseño visual se trabajó en establecer un mapeo de los metadatos para hacer posible la transferencia de información bibliográfica entre el sistema antiguo, *Goobi*, y el nuevo, *Open Access*. En este sentido hay que tener en cuenta que debían migrarse más de 20.000 documentos con más de 1,5 millones de imágenes, cuyo tratamiento técnico había ido variando a lo largo de los años, generando un set de datos heterogéneo. Así, además de los campos bibliográficos tradicionales procedentes de los registros MARC, se crearon una serie de campos que ayudaran a mejorar la navegación y buscabilidad del sistema como colecciones y subcolecciones.

En la siguiente etapa se trabajó arduamente para garantizar que la información estructural, incrustada en los ficheros METS y que enriquece los documentos aportando información de paginación, organización en capítulos, vaciado de láminas e ilustraciones y demás datos con valor añadido, pudiera también traspasarse al nuevo entorno.

En lo tocante a los propios ficheros de imágenes, se evaluó qué formato era el más adecuado para cargar en la nueva herramienta (TIFF, JPG o PDF), optándose finalmente por el TIFF, por ser el que mejor se adaptaba al nuevo visor y aprovechaba al máximo su potencial.

A finales de 2021 comenzaron las pruebas de carga de documentos de distintas tipologías y empezó así una etapa de corrección de errores e incidencias. Se terminó de diseñar las páginas y colecciones, ilustrándose con imágenes adecuadas, y se revisaron y tradujeron los textos descriptivos. Por último, en la etapa final, se hicieron las modificaciones necesarias para asegurar el redireccionamiento correcto desde la *Biblioteca Virtual del CSIC a Simurg*.

Como consecuencia de este proceso de migración y puesta en producción se pueden destacar las mejoras producidas, tanto en la gestión como en la difusión de los objetos digitales. A nivel de gestión interna, disponemos de una herramienta de manejo intuitivo, versátil y personalizable. A nivel de difusión las mejoras han sido realmente destacadas.

En primer lugar, se dispone de visores específicos para cada tipo de material: PDF con OCR para materiales textuales impresos, y TIFF para materiales gráficos y manuscritos.

El visor TIFF para imágenes funciona de tal manera que solo carga las imágenes en alta calidad al hacer zoom sobre ellas. Esto mejora notablemente el rendimiento del portal. Este visor dispone además de un menú de navegación lateral entre páginas que mejora la velocidad de exploración.

Por su parte, el visor para documentos en PDF permite aumentar y reducir las páginas, rotarlas y colocarlas para acomodarse a las necesidades de cada lector. Además, el visor permite buscar en el texto completo de los documentos a través del OCR y navegar entre sus marcadores.

La navegación del portal es intuitiva y sencilla, permitiendo no solo visualizar

los documentos en formato de lista, sino también de manera iconográfica y cronológica a través de una línea de tiempo.

La búsqueda de documentos ha mejorado notablemente y permite recuperar información por todos los campos del registro bibliográfico —se muestren o no en el registro abreviado de *Simurg*—, por el texto completo de los documentos y por los metadatos específicos implementados; esto es, por biblioteca/archivo, por colecciones temáticas, por tipo de documento. El filtrado lateral incluye materias, años, números, autores...

Como novedad significativa respecto a la antigua interfaz, el nuevo portal da la posibilidad de descargar los documentos completos o imagen a imagen, lo que da respuesta a una demanda que se venía produciendo desde hace tiempo.

El nuevo *Simurg* es un buen punto de partida para afrontar con optimismo el futuro de la transformación digital de los fondos patrimoniales del CSIC. Adolece aún de problemas importantes y aún queda la puesta en marcha de otras mejoras que se irán solucionando y generando en los próximos meses.

Por ejemplo, se prevé incorporar la búsqueda en el texto completo de los documentos en el grueso de la colección textual impresa, que aún no dispone de esta funcionalidad. Esto implica un reprocesamiento de los documentos originales que supone una dedicación de recursos muy importante, por lo que se hará paulatinamente y priorizando colecciones determinadas en función de la demanda.

Un tema pendiente es terminar de configurar el sistema para poder volver a ser recolectados por *Europeana*, donde *Simurg* está presente desde 2017. Actualmente se trabaja en adaptar nuestro esquema de metadatos al esquema EDM y se espera implementar el servidor OAI de *Simurg* para seguir colaborando con *Hispana* y *Europeana*.

Por supuesto, se quiere seguir creciendo con las aportaciones de todas las bibliotecas y archivos interesados. En las próximas campañas se hará especial hincapié en incorporar material de los archivos históricos y documentos relacionados con la propia historia del CSIC.

Durante los próximos meses y años se pretende también mejorar la preservación a largo plazo de las colecciones. Desde 2019 se trabaja 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*. Para ello se quiere implantar un software específico de preservación digital, acorde con el modelo OAIS, con la idea de gestionar riesgos, desde los más básicos (fallos en los soportes de almacenamiento), hasta los más complejos, (formatos que quedan obsoletos) o la documentación de los contenidos.

EL OBJETIVO FINAL: DIFUSIÓN Y REUTILIZACIÓN

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

De una parte, desde *Simurg* se apuesta por un uso de las redes sociales que acerque la ciencia a un público no especializado porque, muchas veces, la belleza de la documentación simplemente puede disfrutarse, aunque presente un fin didáctico. Así, se celebran efemérides, días internacionales, eventos de actualidad, las propias novedades del portal o meras curiosidades. Se elaboran productos sobre temas específicos, los llamados *Simurg Temáticos*, donde se agrupa documentación dispersa de diferentes bibliotecas o archivos bajo un hilo conductor. Por último, determinados proyectos merecen, por su envergadura, disponer de un portal propio 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, se incentivan proyectos de reutilización de nuestros contenidos digitales, especialmente en el ámbito de las Humanidades Digitales. En este sentido, por ejemplo, se alienta la participación ciudadana a través del proyecto de descripción colaborativa *Apadrina un documento* (Archivo del Centro de Ciencias Humanas y Sociales, 2021), de la Biblioteca Tomás Navarro Tomás, o se trabaja en la edición digital de los contenidos de *Simurg* en formatos abiertos como el epub.

En esta investigación sobre el propio patrimonio, destaca el portal *Mapping Pliegos*, en el que participa la Biblioteca Tomás Navarro Tomás, 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, se ha colaborado con el grupo Música y Sociedad (IMF-CSIC) en la digitalización de parte del Fondo Reserva de Musicología 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.

REFERENCIAS BIBLIOGRÁFICAS

ARCHIVO DEL CENTRO DE CIENCIAS HUMANAS Y SOCIALES, 2021. *Apadrina un documento*. [en línea]. [Consulta: 1 de septiembre 2022]. Disponible en: <http://biblioteca.cchs.csic.es/DiaLibro21/apadrina.php>

Catálogo Monumental de España (1900-1961). [en línea], 2008. [Consulta: 1 de septiembre 2022]. Disponible en : http://biblioteca.cchs.csic.es/digitalizacion_tnt/

I2S. *Limb Processing*. [en línea], [sin fecha]. [Consulta: 1 de septiembre 2022]. Disponible en: <https://www.limbsuite.com/limb-processing>

INTRANDA. Goobi overview. [en línea], [sin fecha]. [Consulta: 1 de septiembre 2022]. Disponible en: <https://www.intranda.com/en/digiverso/goobi/goobi-overview/>

INSTITUCIÓN MILÁ I FONTANALS DE INVESTIGACIÓN EN HUMANIDADES, 2015. *Arxiu d'Etnografia i Folklore de Catalunya (AEFC)*. [en línea]. [Consulta: 1 de septiembre 2022]. Disponible en: <https://www.imf.csic.es/index.php/servicios/unidad-de-documentacion/fondo-tomas-carreras-i-artau/arxiu-d-etnografia-i-folklore-de-catalunya-aeffc>.

Manuscript@CSIC. [en línea], 2010. [Consulta: 1 de septiembre 2022]. Disponible en: <http://manuscripta.csic.es/>

MORÓN, Isabel e PAREJO, Manuel. El proyecto de digitalización de los

calcos y su finalización. *Enredadera*, n. 27 (2016), p. 20-22. [Consulta: 1 septiembre 2022]. Disponible en: <http://hdl.handle.net/10261/133931>

PONSATI, Agnès, *et al.*, 2010. *Plan Director para la digitalización de fondos del CSIC*. [en línea]. S.l.: [Consulta: 1 septiembre 2022]. Disponible en: <http://hdl.handle.net/10261/23707>

Proyecto de Digitalización de Obras Antiguas de Química y Alquimia. [en línea], 2011. [Consulta: 1 de septiembre 2022]. Disponible en: <http://poq.urici.csic.es/index.html>

SANTAMARINA, Carolina, 2010. *Manual Técnico de Digitalización*. [en línea]. S.l. : [Consulta: 1 de septiembre 2022]. Disponible en: <http://hdl.handle.net/10261/23697>