

EL MUNDO MATEMÁTICO DIGITAL: EL PROYECTO WDML (World Digital Mathematics Library)

Enrique Macías Virgós

*Profesor de Geometría y Topología
Universidad de Santiago de Compostela
Presidente de la Comisión de Información
y Comunicación Electrónicas del CEMAT*



ABSTRACT: We describe the digitization process of the mathematical literature in the framework of the projects WDML and DML-E.

KEY WORDS: Copyright, databases, digitization, mathematical literature, metadata, open access, optical character recognition, scanner.

RESUMEN: Describimos el proceso de digitalización de la literatura matemática de investigación en el marco del proyecto WDML (World Digital Mathematics Library) y del proyecto español DML-E.

PALABRAS CLAVE: acceso abierto, bases de datos, biblioteca digital, derechos de autor, digitalización, escáner, literatura matemática, metadatos, reconocimiento óptico de caracteres.

INTRODUCCIÓN

Los investigadores en matemáticas trabajamos con información y producimos información nueva. Por eso, el acceso a la información es un aspecto crucial de nuestro trabajo. Internet se ha convertido en un gigantesco almacén, con contenidos muy diversos en cuanto a utilidad y calidad, y el desafío es ahora gestionar y hacer accesibles esos contenidos de manera organizada.

Son tres los aspectos que destacaremos: la aparición de las bibliotecas digitales; las innovaciones en publicación electrónica; y el avance de iniciativas como *Open Access* y SPARC¹ [CO].

En el Comité Español de Matemáticas (CEMAT) estamos haciendo un esfuerzo para explicar a nuestros colegas la importancia de estos asuntos que, junto a otros como las bases de datos o los servidores de *preprints*, forman ya parte de la infraestructura básica de nuestra profesión. Para una visión precisa de la situación en España puede consultarse [MA3].

Es típico que los usuarios novatos de bases de datos bibliográficas como *Zentralblatt für Mathematik* o *Mathematical Reviews* se sorprendan de no poder acceder siempre a los textos completos de los trabajos. A continuación explicaré parte del entramado que hay detrás de ese clic que nos permite descargar un artículo en nuestro escritorio.

Agradezco a Jaime Amorós, Pablo de Castro, Manuel González Villa, Juan Luis Varona y Rosa de la Viesca sus observaciones al texto.

DIGITALIZAR

El formato digital significa un cambio enorme en la manera de utilizar el conocimiento. La información electrónica es más accesible y más fácil de reproducir, almacenar y organizar. Además, al menos aparentemente, es más duradera (aunque existe el grave problema de la obsolescencia tecnológica). Diversos autores [GG] han insistido en que, a largo plazo, estos métodos y técnicas tendrán profundas consecuencias económicas, intelectuales y culturales.

WDML

El proyecto de digitalización WDML², auspiciado por la Unión Matemática Internacional IMU, tiene por objetivo último el poner a disposición de todos los usuarios potenciales, a través de Internet, la producción impresa completa de Matemáticas. Y aunque no hay duda de que la biblioteca digital matemática será una realidad en un futuro próximo, las dificultades de llevar esta idea a la

práctica son enormes, como he explicado en un trabajo anterior [MA1].

Además, los conocimientos necesarios están lejos de los usuales en matemáticas: captura de imagen; OCR; protocolos para metadatos; tecnologías de bases de datos; gestión de la propiedad intelectual; diseño de interfaces. Por eso es muy importante la colaboración entre investigadores, bibliotecarios, editores científicos y sociedades matemáticas.

DML-E

Desde 2005 está en marcha el proyecto DML-E de digitalización de todas las revistas de investigación en matemáticas que se publican en España. Este proyecto, auspiciado por el CEMAT en colaboración con el Centro de Información y Documentación Científica CINDOC del CSIC, está financiado por el Ministerio de Educación y Ciencia a través de una Acción Complementaria. Los detalles pueden verse en [MA2]. La responsable del proyecto es Rosa de la Viesca, del CINDOC.

En el marco de WDM existen proyectos similares en otros países, por ejemplo NUMDAM³ en Francia, con quienes mantenemos una estrecha colaboración.

HÁBITOS DE LOS INVESTIGADORES

Según diversos estudios, los investigadores en matemáticas invertimos entre 2 y 10 horas por semana en buscar información. Nuestras fuentes posibles son: las monografías (libros, actas de congresos y tesis); las publicaciones periódicas (revistas, anuarios); las obras de consulta (enciclopedias, diccionarios); las fuentes secundarias (índices, resúmenes, reseñas y bibliografías); los *preprints* y en general toda la "literatura gris" (documentos técnicos, informes, correspondencia). Además, nos gusta hojear las revistas recientes.

Nuestros recursos para buscar información son (por este orden): la biblioteca, los contactos personales o electrónicos con colegas, los congresos, y las consultas a bases de

datos [GV]. Internet, el correo electrónico y la WWW se han convertido también en herramientas importantísimas para nuestro trabajo cotidiano.

Las búsquedas que realiza un matemático en la red son de tres tipos [AZ]: búsquedas bibliográficas (por autor, título, o palabras clave); servicios matemáticos para resolver un determinado problema; y búsquedas de contenidos, por ejemplo artículos que traten de un determinado tema o contengan una determinada fórmula. En el futuro se utilizarán las posibilidades de la web semántica, como en el proyecto Monet [MO], y el objetivo en este campo es desarrollar métodos de formalización del contenido matemático para que sea accesible a los robots, como en el proyecto MoWGLI⁴.

ACCESO ABIERTO

La diseminación del conocimiento no es completa si la información no está accesible. La tendencia actual es que las publicaciones que sean resultado de proyectos de investigación financiados con fondos públicos deben estar accesibles gratuitamente en un plazo corto tras su publicación. Así ha comenzado a implementarse en los EEUU⁵, y en esa línea va la recomendación que publicó en marzo de 2006 la Comisión Europea [D]. Además hay una fuerte presión para que todos los artículos sean de dominio público pasado un período de embargo (usualmente cinco años).

Para difundir estas y otras "buenas prácticas" [C1], que incluyen que cada investigador tenga una página web con sus *preprints* y artículos, es indispensable la ayuda de las sociedades matemáticas.

También algunos autores [DL] han sugerido que el plan de difusión de cada proyecto de investigación comprometa a depositar los resultados de la investigación en un archivo de *preprints* de libre acceso.

Los repositorios institucionales, que van ganando prestigio y visibilidad, sirven para almacenar artículos, prepublicaciones, tesis, actas de congresos y otros documentos de investigación, e incluyen metadatos que son accesibles a través de una interfaz como OAI-PMH⁶ [B].

Además, el movimiento *Open Access* tiene cada vez más adeptos en el terreno de la edición de revistas científicas [ME, GA], debido por una parte al deseo de los científicos por difundir sus resultados y por otra a las restricciones presupuestarias de las bibliotecas, agravadas por los precios impuestos por las editoriales. En el directorio DOAJ⁷ de la Universidad de Lund puede verse una lista actualizada de revistas en acceso abierto.

Citaremos también los directorios OpenDOAR, Scientific Commons, OpenArchives.eu, ROAR y OAlster. Para España, tenemos los portales Scielo Spain⁸, Dialnet y *e-revistas*. El Ministerio de Cultura mantiene un directorio⁹ español de proyectos de digitalización y recursos digitales. También en el portal "Biblioteca Virtual Miguel de Cervantes" hay una lista más o menos actualizada de unos 150 proyectos de digitalización.

DIGITALIZACIÓN

Digitalizar (un libro, un registro sonoro, una película) es transformar en *bits*, en código binario inteligible para un ordenador. Nos interesa aquí sobre todo el paso a formato electrónico de documentos "en soporte papel", especialmente libros y revistas de investigación en matemáticas, pero podrían ser facturas, contratos, manuscritos, mapas, libros de música, periódicos o fotografías.

El resultado de la digitalización (imagen o texto) es un archivo que reproduce el documento original y que puede ser almacenado, indexado, procesado, registrado en una base de datos, y consultado a distancia.

PRESERVAR Y DIFUNDIR

En opinión de muchos expertos [M] "digitalizar no es lo mismo que preservar: la digitalización no es más barata, ni más segura o fiable que la microfilmación". En general, aún se prefiere microfilmarse cuando el objetivo es preservar el material, y digitalizar si el objetivo es facilitar el acceso. Aunque ambos procesos permiten separar la información (el contenido) del soporte o medio físico (y su degradación), la ventaja de la digitalización es la posibilidad de

almacenar múltiples copias en distintos lugares, sin perder información.

PROCESO DE DIGITALIZACIÓN

El proceso de digitalización incluye el escaneado de los documentos y el tratamiento de los datos. Se procede en varias etapas. Primero, se escanea el material para producir una imagen digital. El siguiente paso es el reconocimiento óptico de caracteres (OCR), que nos permite reconocer el texto (y los caracteres matemáticos) para realizar búsquedas. Por último, los "metadatos" deben describir de manera completamente fiable el documento, identificándolo adecuadamente de manera que sirva para darlo a conocer en Internet.

ESCANEADO

Los documentos originales se transforman en imágenes por medio de un escáner, que crea archivos binarios (*raw data*) con muy alta resolución. Estos archivos garantizan que dispongamos de una reproducción fiel de nuestro material. El formato TIFF¹⁰ es el más simple y flexible, aunque cada imagen (una por página) ocupa varios megabytes.

Posteriormente puede reducirse el tamaño mediante alguna técnica matemática de compresión, para generar archivos comprimidos y/o transportables (*pdf, djvu*). En este proceso pueden perderse o no detalles de la imagen original.

Es este último paso el que previsiblemente irá cambiando con los años, y por ello es indispensable que no sólo se guarden los ficheros *.pdf* de cada artículo, sino los *.tif* (las fotos) de cada página, y esto con una resolución buena (mínimo de 600 *dpi*, puntos por pulgada).

Los escáneres que se usan para la digitalización profesional a gran escala son algo diferentes de los planos (de sobremesa) que estamos acostumbrados a ver. Esencialmente son de dos tipos: cenitales, donde el material se digitaliza boca arriba y a distancia; y de tambor, similares a una fotocopiadora industrial, que alcanzan resoluciones de hasta 5000 *dpi*. Para estos últimos, mucho más rápidos, es necesario

desencuadernar cada revista o libro para disponer de hojas sueltas que se introducen en un alimentador de papel. Posteriormente los volúmenes se vuelven a encuadernar. Esto es lo que hacemos en nuestro proyecto DML-E, con lo que el coste de digitalización es relativamente bajo.

Algunos materiales más valiosos, frágiles o que no deben ser desencuadernados se escanean con un escáner cenital, pero es un proceso mucho más lento.

OCR

Tras el escáner, que transforma la imagen en un formato gráfico, suele usarse un programa de reconocimiento óptico de caracteres (OCR), que genera un archivo de texto, que servirá para realizar búsquedas.

Para este fichero se recomienda usar la codificación Unicode, pues en los documentos aparecen muchos caracteres matemáticos o de lenguas distintas del inglés, por lo que no es suficiente el formato ASCII. En ocasiones excepcionales el OCR es sustituido por el teclado del texto original.

Hasta hace poco, el material matemático, como caracteres especiales, fórmulas y diagramas, era difícilmente reconocible por la tecnología OCR, y simplemente se le marcaba en formato gráfico como no reconocido. Esto está cambiando gracias al proyecto "Infty Project" del equipo del Profesor Masakazu Suzuki. Vale la pena hacer una prueba *on-line* de su lector de documentos matemáticos "InftyReader"¹¹, que genera un fichero TeX (el editor de texto mayoritariamente usado por los matemáticos) a partir de una copia gráfica del artículo.

Los programas de OCR cometen un cierto porcentaje de errores, que suelen calibrarse (cuando se van a digitalizar muchos artículos de una misma revista con tipos de letra particulares) sometiéndolos a un proceso de "entrenamiento".

RECOMENDACIONES

Para quien desee acometer un proyecto de digitalización, existen varios manuales y recomendaciones. Todos descri-

ben las técnicas de digitalización, el control de calidad, la gestión de la colección, y la necesidad de que haya un compromiso institucional para preservar a largo plazo el material digitalizado.

Citaremos el tutorial de digitalización editado por la Biblioteca de la Universidad de Cornell [CL], que incluye conceptos básicos e información técnica; las recomendaciones del CEIC [C2] de IMU, diseñadas específicamente para matemáticas en el marco del proyecto WDM; y las directrices de la UNESCO [M], elaboradas por un grupo de expertos de IFLA¹² e ICA¹³, de las que existe una documentada traducción española.

Siempre es necesario aplicar normas internacionales y asumir recomendaciones de buenas prácticas para la captura de imágenes y la creación de metadatos, pues un proyecto de digitalización sólo tendrá éxito si garantiza una interoperabilidad e integración plenas. Por eso es importante evitar el uso de formatos propietarios en los archivos maestros, que en todo caso deben ser de tipo estándar (por ejemplo TIFF, JPEG, ASCII), para facilitar la migración a nuevos formatos que aparecerán en el futuro y para evitar la dependencia de un suministrador particular. Hay ejemplos de dificultades por este motivo, como en el proyecto *Open Book* de la Universidad de Yale.

RECOMENDACIONES DEL CEIC¹⁴

Las recomendaciones específicas que hace el Comité de Información y Comunicación Electrónicas CEIC de IMU¹⁵, y que nosotros hemos respetado en el proyecto DML-E, son las siguientes:

- Resolución de 600 *dpi* (puntos por pulgada), lo que permite discernir pequeños caracteres (subíndices y superíndices, por ejemplo) de tamaño 5 puntos¹⁶
- un alineamiento vertical con un error máximo de 2 grados, para evitar errores en un posterior OCR.

Los ficheros maestros (copia de seguridad de cada página) se guardan sin retoques en formato TIFF *lossless*, es decir comprimidos sin pérdida. Para los archivos de acceso, los algoritmos de compresión eliminan algunos detalles imperceptibles para el ojo humano, Estas imágenes se

agrupan en unidades lógicas (artículo, capítulo de libro, volumen de revista), y se comprimen para su distribución en formato *pdf*.

CONTROL DE CALIDAD

El control de calidad (saltos de páginas, calidad de la imagen) es mucho más estricto que el que se aplica a los microfilmes.

Primero se mejora la calidad de la imagen corrigiendo, mediante programas informáticos, las líneas inclinadas, márgenes inadecuados, fluctuaciones de luminosidad, o manchas y rayas. Después es necesario evaluar la calidad de la imagen obtenida. Por ejemplo, el programa JHOVE¹⁷ examina los ficheros gráficos y valida los detalles de su formato.

Este control de calidad puede hacerse por muestreo aleatorio de las imágenes u otro método.

METADATOS

Toda la información recogida tiene que ser estructurada. Esta gestión documental es la parte más cara y compleja del proceso, por su elevada demanda de esfuerzo de personal.

La manera tradicional de encontrar un documento ha sido a través de los metadatos y los índices. Los metadatos describen un documento, para facilitar la gestión, navegación y uso del material digitalizado. Clásicamente hay tres tipos de metadatos: los descriptivos (identificación); los estructurales (navegación); y los administrativos (gestión). Durante mucho tiempo se usó el formato MARC¹⁸, pero ahora la iniciativa *Dublín Core Metadata* está mejor adaptada a las necesidades de la web.

Como a veces el material que se va a digitalizar carece o tiene errores de catalogación, es indispensable, para el mantenimiento a largo plazo, crear metadatos fiables. En este proceso aparecen términos de difícil traducción, que debemos consultar con expertos. En DML-E nos hemos

topado, por ejemplo, con "variedades catalecticantes", "coyecciones", y con propiedades como "betweenness".

Los metadatos no deben exigir disponer de ningún *software* especial (texto ASCII codificado en XML) y se almacenan en un documento con una DTD¹⁹ disponible públicamente. Como hemos dicho, deben adaptarse al estándar OAI-PMH que se usa para la transmisión de metadatos en Internet.

Es muy importante el método de clasificación que se adopte: los más usados por muchos bibliotecarios son CDU²⁰ y DDC²¹, aunque los matemáticos hemos desarrollado MSC2000²² (ver un estudio completo en [DR]).

El coste de creación de los metadatos es muy alto, pues debe hacerse manualmente y por especialistas en documentación, con asesoramiento de matemáticos. Por eso son de tanto interés distintos proyectos en marcha para generar automáticamente, a partir del texto, la clasificación MSC de un artículo, o sus palabras claves.

DERECHOS DE AUTOR

Los aspectos legales de un proyecto de digitalización se centran sobre todo en los derechos de autor, pues se trata de hacer una copia y una distribución pública, asuntos ambos legalmente protegidos. En la práctica suele pedirse permiso a las revistas, que en general tienen una autorización de los autores para publicar. Para los investigadores de las Universidades y del CSIC se considera que el propietario de los derechos de autor es el organismo que los emplea.

En caso necesario se adoptan fórmulas de *bona fide* que eximen de responsabilidad a los promotores del proyecto, que, por supuesto, es sin ánimo de lucro. Para contratos onerosos, como JSTOR²³, la gestión de los derechos de autor es más compleja.

OBSOLESCENCIA

Nunca hay que perder de vista que la tecnología cambia constante y rápidamente. La obsolescencia tecnológica es un grave problema, que conocemos bien los que hemos



visto pasar discos de vinilo, cintas y casetes, CDs, DVDs y Mp3s. Por no hablar de disquetes, CD-ROMs y memorias flash.

Estos soportes se vuelven rápidamente obsoletos porque la industria de la información y la comunicación no puede esperar por la deseable pero lenta normalización, y esto dificulta las necesidades de archivo a largo plazo. Es conocido, por ejemplo, que algunas versiones antiguas de ficheros *.pdf* ya no son legibles por los programas actuales.

ARCHIVO

Todas las imágenes obtenidas se organizan y nombran de acuerdo con los principios de la archivística, pues es esencial el desarrollo de repositorios digitales fiables basados en la norma OAI²⁴.

Sin embargo, el acceso universal se pervierte si la información no es estable (por ejemplo, debido a la existencia de múltiples versiones de un documento) o no tiene un URL²⁵ bien definido.

Curiosamente, la identificación estable de un documento no garantiza su autenticidad. En ocasiones se han detectado cambios deliberados o fraudulentos en el contenido de los ficheros (por ejemplo a causa de un conflicto laboral). Conviene adoptar métodos para detectar estos posibles cambios, así como para marcar la propiedad de las copias digitalizadas, aunque se las distribuya gratuitamente en Internet.

Por último, hay que almacenar las copias en lugar seguro. Una buena estrategia es distribuir las copias a las revistas y otros repositorios, con lo que la localización física del documento se vuelve irrelevante.

TEXDOC

Una característica propia de las matemáticas es el uso del editor de texto TeX (o su dialecto LaTeX), que es aceptado o exigido por la mayor parte de las editoriales de libros y revistas. En los últimos años TeX se ha per-

feccionado y ahora incluye posibilidades como inclusión de gráficos (por ejemplo en PostScript); creación de documentos en formato pdf; uso de hiperenlaces internos y externos; y color. También hay conversores de TeX a MathML.

En el *TeX Document Center*²⁶ es posible certificar que el código TeX que hemos usado en nuestros artículos se ajusta a los estándares, y que lleva los macros y ficheros de estilo necesarios para su compilación futura.

En matemáticas sería deseable preservar el fichero TeX de cada artículo (es un fichero de texto en formato ASCII).

PROYECTOS EN MARCHA

En Matemáticas, los proyectos de digitalización más avanzados son JSTOR (Cornell University, Ithaca, EEUU), SUB (Göttingen, Alemania) y NUMDAM (Grenoble, Francia). El proyecto global WDML es gestionado por IMU. Para más detalles consultar [J].

La Unión Europea tiene distintas iniciativas relacionadas con la digitalización, el acceso abierto y las bases de datos internacionales. Del 5.º programa marco citaremos solamente LIMES²⁷, en el que hemos participado, para potenciar *Zentralblatt für Mathematik*; y EULER²⁸, un portal que permite acceder a catálogos y repositorios de documentos matemáticos. A partir de 2003, la Comisión decidió apoyar solamente proyectos cofinanciados con organizaciones comerciales, y en el 6.º programa marco se rechazaron todos los proyectos de matemáticas, incluidos DML-EU (digitalización) y CITIZEMS (bases de datos) presentados por la Sociedad Matemática Europea EMS.

En marzo de 2005 la Comisión Europea puso en marcha el programa *e-Content plus*, financiando el proyecto EDL²⁹, gestionado por las Bibliotecas Nacionales de varios países, con el objetivo de coordinar sus catálogos y facilitar el acceso a sus colecciones digitales.

Recientemente, la iniciativa europea *i2010 Digital Libraries Initiative* pretende poner en la red todo el material científico y cultural europeo, en un entorno multilingüe.

OTROS PROYECTOS

Hay otras posibilidades interesantes de digitalización que han sido poco estudiadas y catalogadas, como la llamada "literatura gris" (informes, boletines, documentos técnicos, trabajos no publicados). Por ejemplo, Brian Davies es el impulsor de un repositorio de apuntes y material docente que se llama *Mathematics Online*. En España tenemos el repositorio de apuntes en castellano del CEMAT.

TESIS DOCTORALES

Como ha señalado Andrés Pedreño, ex-consejero delegado de Universia, "las universidades tienen una inmensa responsabilidad en la tarea de impulsar la digitalización del conocimiento. Libros, tesis, publicaciones... no pueden quedar en formato de papel, tras costosas inversiones públicas, en las estanterías de nuestras Bibliotecas".

Las tesis doctorales son textos de extraordinario valor académico, pero muy especializados, por tanto sin demanda comercial. Sin embargo, puede aprovecharse el potencial de difusión de Internet. Muy recientemente, DOCUMAT-RSME, con la colaboración de DIALNET y el CINDOC, ha lanzado una iniciativa para digitalizar y dar acceso a las tesis doctorales de matemáticas leídas en España.

TESEO es la base de datos que recoge, desde 1976, todas las tesis doctorales leídas y calificadas favorablemente en las universidades españolas, Depende del Consejo de Coordinación Universitaria, y la información procede directamente de las Comisiones de Doctorado de las universidades. Cada tesis tiene un código de referencia único, para facilitar las búsquedas.

Hay también un proyecto internacional (basado en Alemania) que se llama *MathDiss*. La UNESCO elaboró hace tiempo unas recomendaciones sobre la edición electrónica de tesis en el marco del proyecto ETD³⁰.

PORTALES

Para dar acceso a los materiales digitalizados es necesario crear un "portal". Un portal temático puede tratar sobre temas tan diversos como prensa histórica, autores gallegos,

grabaciones de voces, Latinoamérica, patrimonio musical o cartografía.

En matemáticas, el primero fue ERCIM³¹, pero el más famoso sin duda es arXiv, que es un servidor electrónico para distribuir artículos (normalmente *preprints*). Creado por el físico Paul Ginsparg en Los Álamos, desde 2001 está alojado en Cornell. ArXiv da acceso a más de 400.000 *e-prints* en física, matemáticas, computación y biología cuantitativa.

Citaremos también MPRESS, al que se accede desde EMIS³², y mp_arc, creado en 1991 por Rafael de la Llave en Austin, Texas. Otros portales de matemáticas son por ejemplo Intute, en el Reino Unido; y EMANI³³ que es una iniciativa conjunta de las bibliotecas de las Universidades de Cornell (EEUU), Tsinghua (Pekín, China), Göttingen (Alemania) y Orsay (París) junto con editoriales como Springer y Birkhäuser.

EMIS

EMIS es el servidor de información de la Sociedad Matemática Europea (EMS), en colaboración con la fundación FIZ Karlsruhe. Sus principales componentes son [WJ]: la biblioteca electrónica *ElibM*³⁴, y las bases de datos ZMATH (1931-) y JFM³⁵ (1868-1942). Tiene unos cuarenta "espejos" en todo el mundo.

La "Biblioteca Electrónica de Matemáticas" *Elibm* da acceso, casi siempre libre, al contenido de más de 60 revistas de matemáticas (también actas de congresos y monografías), tanto estrictamente electrónicas como versiones electrónicas de revistas impresas. La decisión de incluir o no una nueva revista se toma de acuerdo con el Comité de Publicaciones Electrónicas de la EMS, del que forma parte el autor de este artículo.

RECURSOS DE INTERÉS

Entre los recursos de interés en el mundo de la información y documentación citaremos DOIS³⁶, en el que pueden descargarse unos 14.000 documentos y que es "el portal

internacional de referencia para el profesional de la información" [AR].

También es excelente el servidor internacional de *e-prints* E-LIS³⁷, especializado en información y documentación científicas, que gestiona en Italia Antonella de Robbio. Es un archivo no comercial, que se nutre de las contribuciones de los propios investigadores, que tienen así la oportunidad de archivar sus trabajos y darles una gran visibilidad.

Una herramienta interesantísima es el catálogo colectivo de la Red de Bibliotecas Universitarias REBIUN, que permite consultar los fondos de 67 bibliotecas universitarias y de investigación españolas.

CEMAT

En la página web del CEMAT también mantenemos, gracias a los desvelos de Manuel González Villa, una lista actualizada de artículos y noticias sobre publicaciones electrónicas.

CONCLUSIÓN

Las razones para convertir documentos al formato digital son, en primer lugar, la necesidad de abaratar y mejorar el acceso, en facilidad o en calidad, y en segundo lugar reducir el uso de los materiales originales para evitar su deterioro (especialmente si son frágiles), creando una copia de seguridad.

Al diseñar un proyecto de digitalización debemos estudiar [M]: su viabilidad técnica; los aspectos legales; los criterios

de selección; cómo se va a acceder a los materiales; definir las normas de escaneado y de generación de metadatos; cómo preservar y dar valor añadido al material generado. Además, debe preverse la creación de un punto de acceso, y hay que tener una política clara de acceso y uso de las imágenes.

Los costes del proyecto deben asegurar el control de calidad y la incorporación de metadatos, e incluirán el tiempo del personal especializado; el escaneado; el almacenamiento en disco; el mantenimiento; la migración a nuevos soportes y formatos. También el precio de encuadernar y reencuadernar.

Para seleccionar el material, debemos conocer los intereses de los usuarios. Y es necesario implicar a especialistas e investigadores en la toma de decisiones.

En nuestro caso, el equipo del proyecto DML-E incluye especialistas del CINDOC, matemáticos del CEMAT y editores de revistas. Hemos conseguido interesar a todas las revistas de investigación de matemáticas que se publican en España. Y hemos digitalizado ya unas cien mil páginas.

Para un organismo como el CINDOC, cada proyecto de este tipo es una oportunidad para desarrollar aun más su infraestructura técnica, incrementar sus servicios, formar a su personal, establecer colaboraciones con otras instituciones y extender su influencia.

Y para la comunidad matemática española significa demostrar que estamos al nivel de los países más avanzados dentro del proyecto WDML de digitalización de toda la literatura matemática, para crear la futura Biblioteca Digital Mundial.

NOTAS

- 1 *Scholarly Publishing and Academic Resources Coalition.*
- 2 *World Digital Mathematics Library.*
- 3 *Numérisation de documents anciens mathématiques.*
- 4 *Mathematics on the Web: Get It by Logics and Interfaces.*
- 5 *NIH Public Access policy*, en vigor; *Federal Research Public Access Act of 2006*, propuesta.
- 6 *Open Archives Initiative - Protocol for Metadata Harvesting.*
- 7 *Directory of Open Access Journals.*
- 8 *Scientific Electronic Library online.*
- 9 <http://roai.mcu.es/es/inicio/inicio.cmd>
- 10 *Tagged Image File Format.*
- 11 http://www.inftyproject.org/en/OnlineService_E.html
- 12 *International Federation of Library Associations and Institutions.*
- 13 *International Council on Archives.*
- 14 *Committee on Electronic Information and Communication.*
- 15 *International Mathematical Union.*
- 16 *Una pulgada (25,2 mm) contiene 12 picas o sea 72 puntos.*
- 17 *JSTOR/Harvard Object Validation Environment.*
- 18 *Machine Readable Cataloging.*
- 19 *Document Type Definitions.*
- 20 *Clasificación Decimal Universal.*
- 21 *Dewey Decimal Classification.*
- 22 *Mathematics Subject Classification.*
- 23 *Journal Storage: the scholarly journal archive.*
- 24 *Open Archive Information System.*
- 25 *Uniform Resource Locator*, la dirección en Internet.
- 26 <http://www.texdocc.org/>
- 27 *Large Infrastructure in Mathematics. Enhanced Services.*
- 28 *European Libraries and Electronic Resources in Mathematical Sciences.*
- 29 *European Digital Library.*
- 30 *Electronic Theses and Dissertations.*

- 31 *European Research Consortium for Informatics and Mathematics.*
- 32 *European Mathematics Information Service.*
- 33 *The Electronic Mathematics Archiving Network Initiative.*
- 34 *Electronic Library of Mathematics.*
- 35 *Jahrbuch über die Fortschritte der Mathematik.*
- 36 *Documents in Information Science.*
- 37 *Library and Information Profession Eprints Server.*

BIBLIOGRAFÍA

- [AR] Alonso Arévalo, Julio. "Documents in Information Science" (DoIS): portal internacional de referencia para el profesional de la Información. En *Proceedings Congreso Internacional de Información INFO 2004*, La Habana (Cuba), 2004. <http://eprints.rclis.org/archive/00001367/>
- [AZ] Asperti, Andrea; Zacchiroli, Stefano. *Searching Mathematics on the Web: State of the Art and Future Developments New Developments in Electronic Publishing AMS/SMM Special Session*, Houston, May 2004 ECM4 Satellite Conference, Stockholm, June 2004, pp. 9-18. <http://www.emis.de/proceedings/Stockholm2004/asperti.pdf>
- [B] Barrueco Cruz, José Manuel and Subirats Coll, Imma. Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH): descripción, funciones y aplicación de un protocolo. *El Profesional de la Información* 12(2), pp. 99-106 (2003). <http://eprints.rclis.org/archive/00000177/>
- [C1] CEIC-IMU Recomendaciones sobre Información y Comunicación. Comité de Información y Comunicación Electrónicas de la Unión Matemática Internacional (IMU). *La Gaceta de la RSME*, Vol. 6.3 (2003), pp. 617-625.

Recibido: 10 de enero de 2007

Aceptado: 17 de enero de 2007

- [C2] CEIC-IMU. *Some Best Practices for Retrodigitization*. Endorsed at Santiago de Compostela by the General Assembly of the International Mathematical Union, August 20, 2006. http://www.ceic.math.ca/Publications/retro_bestpractices.pdf
- [CL] Tutorial de digitalización de imágenes (versión en español), Departamento de investigación, Biblioteca de la Universidad de Cornell, 2000-2003. <http://www.library.cornell.edu/preservation/tutorial-spanish/contents.html>
- [CO] Comba, Valentina. Trends in digital libraries and scholarly communication among European Academic Research Libraries. *California Digital Libraries Discussion Group*, Irvine, January 15th, 2004. <http://eprints.rclis.org/archive/00000758/>
- [D] Dewatripont, Mathias *et al.* *Study on the economic and technical evolution of the scientific publication markets in Europe*, Final Report. European Commission. Directorate-General for Research, 2006. http://ec.europa.eu/research/science-society/pdf/scientific-publication-study_en.pdf
- [DL] De la Llave, Rafael. La publicación electrónica de la investigación matemática: Apuntes para un debate sobre su economía y sociología, *La Gaceta de la RSME*, Volumen 6, No. 2 (mayo-agosto, 2003).
- [DR] De Robbio, Antonella; Maguolo, Dario; Marini, Alberto. Mathematics Subject Classification (MSC) and related Schemes in the Digital World. En *Proceedings Eighth International Conference "Crimea 2001" "Libraries and Associations in the Transient World: New Technologies and New Forms of cooperation"* 1.(Tom 1.), June 9-17, 2001. <http://eprints.rclis.org/archive/00000022/>
- [GA] Gómez, Nancy D.; Arias, Olga M. El movimiento de acceso abierto y los repositorios institucionales. *3a. Jornada sobre la Biblioteca Digital Universitaria*, Cordoba, Argentina (2005). <http://eprints.rclis.org/archive/00007785/>
- [GG] González Quirós, José Luis; Gherab Martín, Karim: *El templo del saber. Hacia la biblioteca digital universal*. Ediciones Deusto, 2006.
- [GV] Guevara Villanueva, Angélica: El comportamiento informativo de los investigadores en el área de las matemáticas, *Liber, Revista de Bibliotecología*, Nueva época, Vol. 6, Núm. 1, enero-marzo 2004.
- [J] Jackson, Allyn. The digital mathematics library. *Notices of the American Mathematical Society*, September 2003. <http://www.ams.org/notices/200308/comm-jackson.pdf>
- [MA1] Macías Virgós, Enrique. Un gran proyecto de cooperación internacional: la Biblioteca Digital de Matemáticas. *La Gaceta de la RSME* 6, 2, 351-366 (2003).
- [MA2] Macías Virgós, Enrique. Some digitization initiatives in Spain. En *New Developments in Electronic Publishing of Mathematics*, ECM4 Satellite Conference, Stockholm, June 2004. Becker, H.; Stange, K.; Wegner B. Eds. FIZ Karlsruhe 2005, pp. 137-142. <http://www.emis.de/proceedings/Stockholm2004/>
- [MA3] Macías Virgós, Enrique. The Spanish DML-E digitization project and related topics, CMDE 2006 Communicating Mathematics in the Digital Era, Aveiro, August 15-18 2006. http://xtsunxet.usc.es/macias/material/aveiro_2006.pdf
- [MA4] Macías Virgós, Enrique. La importancia de las bases de datos en Matemáticas, en *Análisis y perspectivas de la colaboración latinoamericana en Matemáticas*, Publ. de la RSME, pp. 18-24 (2003). <http://web.usc.es/~xtquique/OtherPapers.html>
- [M] Macllwaine, John *et al.* Directrices para proyectos de digitalización de colecciones y fondos de dominio público, en particular para aquéllos custodiados en bibliotecas y archivos. IFLA; ICA. UNESCO. Traducción al castellano por A. I. Cámara *et al.*, 121 págs. Ministerio de Cultura, 2002. http://travesia.mcu.es/documentos/pautas_digitalizacion.pdf
- [ME] Melero, Remedios (2005) Acceso abierto a las publicaciones científicas: definición, recursos, copyright e impacto. *El Profesional de la Información* 15(4), pp. 255-66. <http://eprints.rclis.org/archive/00004371/>
- [MO] Mathematical Service Description Language: The MONET Consortium (IST-2001-34145). Deliverable D14. <http://monet.nag.co.uk/cocoon/monet/publicdocs/monet-msdl-final.pdf>
- [WJ] Wegner, Bernd; Jost, Michael. EMIS: A Portal to Mathematics in Progress. *High Energy Physics Libraries Webzine* (6/2002). <http://www.rsme.es/pe/wegner.PDF>

**Enlaces de interés
[revisados el 3/1/2007]**

arXiv
<http://arxiv.org/>

Base de datos TESEO de tesis españolas
<http://www.mcu.es/TESEO/>

Biblioteca Virtual Miguel de Cervantes
<http://www.cervantesvirtual.com/>

Catálogo colectivo de REBIUN
<http://rebiun.crue.org/>

Comité Español de Matemáticas CEMAT
<http://www.ce-mat.org>

Comité de Información y Comunicación
Electrónicas IMU
<http://www.ceic.math.ca/>

Dialnet
<http://dialnet.unirioja.es/>

DOAJ
<http://www.doaj.org>

DOIS
<http://wotan.liu.edu/doi/>

EDL
<http://www.edlproject.eu>

E-LIS
<http://eprints.rclis.org/>

EMIS
<http://www.emis.de/>

e-revistas
<http://www.tecnociencia.es>

Infty Project
<http://www.inftyproject.org/en/index.html>

INTUTE Science,
Engineering and Technology
<http://www.intute.ac.uk/sciences/mathematics/>

JSTOR
<http://www.jstor.org/>

MathDiss
<http://www.ub.uni-duisburg.de/mathdiss/>

Mathematics OnLine Open Access Lecture
Notes and Monographs
<http://www.mth.kcl.ac.uk/MAO/>

MoWGLI
<http://mowgli.cs.unibo.it/>

mp_arc
http://www.ma.utexas.edu/mp_arc/

NUMDAM
<http://www.numdam.org/>

OAI_PMH
[http://www.openarchives.org/OAI/2.0/
openarchivesprotocol.htm](http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm)

REBIUN
<http://rebiun.absysnet.com/>

SCIELO
<http://www.scielo.org/index.php?lang=en>

SPARC
<http://www.arl.org/sparc>

WDML World Digital Mathematics Library
<http://www.ceic.math.ca/WDML/index.shtml>

