

Un gran proyecto de cooperación internacional: la Biblioteca Digital de Matemáticas

por

Enrique Macias–Virgós

INTRODUCCIÓN

Desde hace mucho tiempo, la actividad profesional de los matemáticos depende de las búsquedas y consultas bibliográficas. Durante casi 2500 años, las publicaciones matemáticas esenciales han ido pasando de generación en generación. Con las nuevas herramientas electrónicas nos hemos habituado a disponer, de manera casi instantánea, de material, información y datos sobre nuestra mesa de trabajo. Y sin embargo, los fondos disponibles en formato digital son sólo una pequeña parte del enorme patrimonio cultural y científico de las Matemáticas.

Con el inicio del siglo XXI, una ambiciosa idea ha ido tomando cuerpo: digitalizar toda la literatura matemática de todos los tiempos, para crear una gigantesca biblioteca virtual a disposición de la comunidad matemática internacional. Un enorme proyecto de cooperación a escala planetaria, con una duración estimada de 10 a 15 años, y un coste de más de 100 millones de euros, para convertir 50 millones de páginas a formato electrónico, accesibles desde cualquier lugar del mundo.

En este artículo, que es un resumen de distintos documentos que la RSME quiere dar a conocer a la comunidad matemática española, contamos las ventajas y dificultades del proyecto DML (*Digital Mathematical Library* o Biblioteca Digital de Matemáticas), los pasos que se han dado, y la participación de la RSME en el equipo europeo que lo está preparando.

1. LITERATURA MATEMÁTICA: LA SITUACIÓN ACTUAL

Las Matemáticas se distinguen de otras Ciencias por su dependencia de las publicaciones antiguas y por la longevidad de sus resultados y métodos. En primer lugar [21], las referencias interesantes no siempre son las más recientes porque los trabajos de investigación de calidad nunca quedan obsoletos, debido a nuestra tradición de no repetir demostraciones en los nuevos artículos, sino citar la fuente original, que muchas veces puede haber sido publicada hace 30 o 40 años. Por otra parte, muchas ideas y técnicas para resolver problemas actuales proceden de trabajos publicados hace mucho tiempo, por lo que esos artículos antiguos no solamente tienen un considerable interés histórico

sino que inspiran o sirven de base a resultados novedosos en investigación. Por ejemplo [14], en los *Annals of Mathematics* de Princeton los artículos del año 2001 contenían un 60% de referencias a trabajos anteriores a 1992; recíprocamente, en los artículos de otras revistas que en 2001 citaban algún trabajo de los *Annals*, el 82% de las citas se refería a trabajos anteriores a 1992. Se han obtenido resultados análogos para el *Inventiones Mathematicae* y para el *J. Reine und Ang. Math.* Basta ver además el *Science Citation Index* (<http://www.isinet.com/isi/products/citation/sci/>), que publica datos estadísticos de las mejores revistas: la vida media de los artículos de Matemáticas es muchas veces de más de 10 años (el máximo posible), mientras en otras áreas raramente supera los 6 meses.

BASES DE DATOS [22] La literatura matemática está extraordinariamente bien estructurada en comparación con el corpus bibliográfico de cualquier otra disciplina, lo que facilita la idea de archivarla y usar herramientas automáticas de búsqueda. En parte esto se debe a la existencia de excelentes bases de datos y servicios de recensiones como *Zentralblatt MATH* (de la EMS) y *Mathematical Reviews* (de la AMS), que son accesibles *on-line* y han servido para organizar lo que sin ellas sería *un montón de revistas irremediablemente desorganizado* [20].

Como todas las Ciencias, las Matemáticas crecen a un ritmo cada vez mayor. Por ejemplo, en la base de datos ZMATH (<http://zmath.cesga.es>), que publica recensiones de artículos desde 1931, el número total de trabajos catalogados supera ya 1.800.000, a un ritmo aproximado de 80.000 nuevos artículos por año.

Hay una explosión de material matemático *on-line* (individuos, editoriales, organismos...), y la comunicación ha mejorado enormemente con herramientas como editores de texto (TEX) o archivos de *preprints* y artículos (*arXiv*). También aumenta sin cesar el número de revistas publicadas en formato electrónico.

Sin embargo, la cantidad de documentos disponibles en formato digital es proporcionalmente muy pequeña. Por eso, a medida que hay más literatura matemática accesible en Internet, más se echa de menos que no exista la misma posibilidad de hacerlo con los trabajos antiguos.

BIBLIOTECAS [12] Los consumidores de Matemáticas (matemáticos, ingenieros, científicos, educadores, historiadores...) necesitan acceder a las publicaciones de forma rápida, fiable e integrada, y lo ideal sería poder disponer de la colección completa de todas las publicaciones actuales y antiguas. De hecho, hasta hace unos años las grandes universidades y centros de investigación solían disponer de bibliotecas de Matemáticas muy completas, con las mejores revistas en formato papel, y esto cubría satisfactoriamente las necesidades de los usuarios de su zona de influencia.

Pero esta política se ha hecho insostenible debido a los precios cada vez mayores que fijan las editoriales y al incremento exponencial del número de

publicaciones. Podemos hacernos una idea con un ejemplo, tomado entre otros varios en España: la Facultad de Matemáticas de la Univ. de Santiago, con algo menos de 100 profesores y 1000 alumnos, recibe, a través de suscripciones o intercambios, del orden de 200 revistas (cada una con varios números al año), y gasta un presupuesto anual de más de 250.000 euros (40 millones de pesetas) en suscripciones a revistas, y una décima parte (25.000 euros) para compra de libros.

Por otra parte, en los centros de los países más desarrollados es habitual el acceso en red a catálogos electrónicos y bases de datos bibliográficos. Pero no está generalizado el acceso a las versiones completas de los documentos en formato electrónico, pese a que la publicación de un artículo se hace ahora siempre a través de tratamientos de texto y los autores deben enviar a las editoriales la versión digital de su trabajo. Además, y por diferentes motivos, en muchas bibliotecas hay grandes lagunas en las colecciones de revistas.

La situación cambia drásticamente en los países menos desarrollados, o en las universidades más pequeñas, donde el acceso a la literatura matemática es difícil o inexistente, y las bibliotecas carecen de medios.

2. PRELIMINARES DEL PROYECTO: IDEAS BÁSICAS

En el año 2000, Philippe Tondeur, conocido geómetra de la Univ. de Illinois y entonces Director de la División de Matemáticas de la NSF (*National Science Foundation*, EEUU), lanzó la idea de un proyecto internacional de digitalización masiva para crear la Biblioteca Virtual Mundial de Matemáticas, *la versión actualizada del sueño de la biblioteca de Alejandría* [25] y hacer accesible a través de Internet la literatura matemática, para todos los matemáticos del mundo.

El proyecto DML (*Digital Mathematics Library*) persiguiría tres objetivos globales:

1. crear una colección representativa de la mayor parte de la literatura matemática, digitalizando tanto documentos antiguos como publicaciones recientes;
2. adoptar estándares técnicos adecuados a las nuevas tecnologías de la información y la comunicación (TIC);
3. establecer los mecanismos necesarios para preservar la colección, mantenerla permanentemente actualizada, y ponerla a libre disposición hasta donde lo permitan las normas de *copyright*.



Philippe Tondeur, propuso el proyecto DML

En el primer artículo que apareció tratando de concretar estas ideas, y que es una referencia imprescindible [10], John Ewing, director ejecutivo de la *American Mathematical Society* (AMS), escribió: *Es difícil imaginar otro cambio que tenga tanta influencia. El proyecto tendrá un gran impacto en la forma en que se hacen y usan las Matemáticas, y servirá probablemente como prototipo para una nueva forma de cooperación científica en la sociedad electrónica del siglo XXI.*

CALENDARIO Se trataría de un proyecto con una duración estimada de entre diez y quince años, estructurado en tres fases:

1. concepción y diseño;
2. digitalización;
3. archivo y distribución.

La primera fase duraría dos o tres años. En ella habría que crear unas normas básicas de digitalización y decidir qué protocolos y estándares técnicos se van a utilizar, para que la literatura sea accesible tanto en la actualidad como en el futuro. Habría también que negociar acuerdos sobre derechos de autor; decidir qué material se va a incluir; establecer cómo se reparte el trabajo entre diferentes grupos; coordinar el archivo del material en una red formada por muchos servidores y *espejos* en diferentes países; y establecer cuáles serán los procedimientos para agregar material en el futuro.

En la segunda fase, de implementación, comenzaría el trabajo de digitalización masiva propiamente dicho. Por digitalizar se entiende no sólo crear imágenes escaneadas de artículos y libros, susceptibles de ser examinadas por un programa de reconocimiento óptico de caracteres (OCR), sino que hay que incluir también un fichero de texto asociado (que haga posible disponer de servicios de búsquedas de todo tipo), y los llamados *metadatos* (que permiten la inclusión de enlaces electrónicos y datos bibliográficos de identificación).

Para terminar, habría una tercera fase de puesta en marcha y mantenimiento a largo plazo de un sistema de distribución del contenido de los archivos, probablemente a través de una red de bibliotecas de referencia, que lo pondrían a disposición de distintas organizaciones, bibliotecas universitarias y centros de documentación, para diseminarlo entre la comunidad matemática internacional mediante conexiones en red y licencias de acceso.

3. DIFICULTADES DEL PROYECTO

Un proyecto tan ambicioso plantea desafíos enormes, y aunque la idea básica es clara, hay muchos obstáculos difíciles de salvar.

ORGANIZACIÓN En primer lugar tenemos dificultades de carácter organizativo. El proyecto será llevado a cabo a través de la cooperación internacional (EEUU, Europa, Rusia, China...) y todas las sociedades matemáticas del mundo lo apoyan (al menos moralmente). Pero hay que decidir quién participa, de qué se encarga cada país y qué organismo llevará a cabo las tareas de coordinación. Por ejemplo, hay serias dudas de que la IMU (Unión Matemática Internacional) tenga capacidad para asumir esta responsabilidad, ya que carece de personalidad jurídica.

Por otra parte [11] es esencial que el proyecto no sea visto como propiedad de un único grupo, ni que sea llevado a cabo por una empresa privada con ánimo de lucro. Por eso interesa que haya el mayor número posible de participantes (incluso aunque no puedan contribuir financieramente) y que estén representadas todas las partes afectadas (editoriales, bibliotecas, usuarios...) para combinar sus intereses y optimizar el esfuerzo económico.

En principio, un equipo científico, técnico y de consultoría actuaría como centralizador y administrador del proyecto, estableciendo un control de calidad del trabajo bajo la supervisión de representantes de la comunidad matemática que den su opinión sobre las decisiones más importantes que haya que tomar. En la actualidad hay diferentes comités DML de coordinación tanto con IMU como con los socios europeos y norteamericanos, así como distintos grupos de trabajo sobre estándares técnicos, archivo y modelo económico del proyecto.

A su vez, DML dirigiría el trabajo de muchos grupos y comités en diferentes países, y coordinaría distintos proyectos nacionales de digitalización, financiados por los gobiernos respectivos e involucrando a varias sociedades, bibliotecas, editoriales y centros de documentación. El objetivo en este caso es digitalizar un bloque dado de literatura (determinadas revistas, editoriales o épocas...), trabajo que puede ser llevado a cabo por subcontratistas, pero asegurando que DML guarda copia de todo lo que se digitalice con fondos públicos.

CONTENIDOS Otra dificultad es la falta de una definición precisa de qué entendemos por *literatura matemática*. A los cientos o miles de revistas de investigación y didáctica, de calidad desigual, que se editan en la actualidad, hay que añadir las que publican artículos de Matemáticas mezclados con otras ciencias. La solución de incluir solamente colecciones completas de ciertas revistas dejaría por tanto fuera una gran cantidad de artículos de interés; por otra parte, el ir haciendo una selección caso por caso de artículos de una revista dada añadiría complicaciones y costes enormes.

Como hemos dicho, los artículos de investigación en Matemáticas son por regla general *pequeños*, en el sentido de que contienen todas las referencias necesarias para entender completamente su contenido. Sin embargo, los textos más antiguos (por ejemplo, griegos, indios, chinos, bizantinos o árabes) plantean problemas muy distintos: desde el formato, hasta el hecho frecuente de que están incluidos en contextos no matemáticos y por tanto deben ser

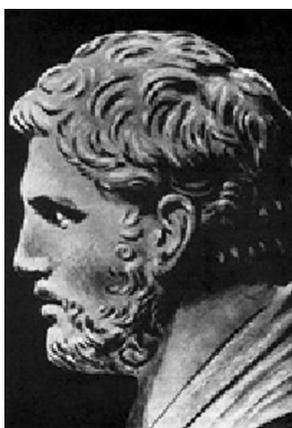
aislados y volcados fuera del resto del texto, probablemente acompañados de un resumen apropiado del resto del libro y de otros textos con comentarios, interpretaciones o valoraciones que destaquen su importancia, autenticidad o utilidad, e información bibliográfica e histórica sobre el autor.

La situación con los libros es aún peor. Deberá decidirse si se incluyen o no los libros de texto, de qué nivel, con qué contenidos no estrictamente matemáticos sino en la frontera con otras áreas, o cuál de las varias ediciones de una misma obra.

Otro problema esencial es decidir qué idiomas se tendrán en cuenta.

ESTÁNDARES TÉCNICOS ¿Cómo crear estándares e interfaces para Matemáticas que sean compatibles con los de otras disciplinas y con los que puedan establecerse para toda la web? Los grupos de trabajo de DML han presentado ya unas primeras recomendaciones [3]. Se refieren a la calidad del escaneo (resolución mínima, normas ISO); a los formatos de archivo (tiff, ZIP...); los nombres de ficheros (URL, MSC 2000...); los formatos de entrega de los documentos (pdf, djvu...); los enlaces a bases de datos (MR, Zbl...); el reconocimiento óptico de caracteres (OCR) y el reconocimiento de fórmulas; las unidades mínimas de digitalización (artículos, capítulos de libros, tablas de contenidos...); las técnicas de servidores (apache); los mecanismos de conversión de formatos, etc.

Una posibilidad es crear una red de servidores, dotados de buena infraestructura, donde cualquiera pudiese enviar sus artículos en papel para escanearlos con alta calidad, y proporcionar herramientas públicas para transformar el material al formato correcto. Ya existen prototipos de estos servidores [26].



Demetrio de Falero (350 A.C.–283), impulsor de la biblioteca de Alejandría

ARCHIVO DEL MATERIAL El problema de archivar y catalogar tal cantidad ingente de material es interesante no sólo desde el punto de vista de las ciencias de la computación sino de los matemáticos teóricos. Por ejemplo, en [22] se sugiere una estructura de espacio métrico (!) para la Biblioteca Virtual, en cuyo centro estarían los trabajos de Arquímedes y Euclides, y en la corteza los *preprints*, esperando en cola ser aceptados por alguna revista... En este aspecto, el proyecto es una excelente oportunidad para áreas como diseño de arquitecturas técnicas de la información, desarrollo de sistemas, minería de datos y modelado matemático del espacio de información [24].

Por *archivo* se entiende preservación y acceso sostenible a largo plazo. Pero, a

diferencia del material impreso, los esfuerzos para mantener ficheros digitales son inconsistentes y fragmentarios. Aunque existen estándares ISO (se ha desarrollado un modelo de referencia llamado OAIS *Open Archival Information System* [17]), no hay todavía reglas generalmente aceptadas para preservar información digital, y de hecho mucho material electrónico se ha perdido por falta de una planificación a largo plazo.

Hay cierto debate [18] sobre si los documentos deben conservarse en su formato original o convertirlos a unos pocos formatos bien documentados. Lo primero permite a los futuros investigadores apreciar mejor las características del documento; lo segundo facilita la conversión en el futuro para asegurar su legibilidad. Otro problema es si el material del depósito queda ahí para siempre, o si deben hacerse revisiones periódicas a medida que cambian las prioridades y criterios de selección. En opinión de varios expertos [10], modificar el formato de los archivos es una tarea muy delicada, y es casi imposible prever qué datos deben ser conservados en cada momento para que el trabajo no se pierda.

COPYRIGHT La legislación sobre derechos de autor es extremadamente compleja, y diferente en cada país. Se trata de negociar no sólo con cientos de editoriales diferentes, sino que en muchos casos ni siquiera estará claro a quién pedir los permisos correspondientes: autores ya fallecidos, empresas desaparecidas, materiales de origen incierto o desconocido. Estos problemas requieren disponer de un equipo jurídico. Además, a algunas editoriales les disgusta la idea de un portal común de acceso, y creen que es suficiente con las bases de datos MR y ZMATH; para otras sin embargo, el añadir material a la colección se convertirá en una parte más del proceso de publicación, aprovechando las ventajas de la estandarización.

En cualquier caso, parece ir imponiéndose la noción de *moving wall*, es decir el acuerdo para que tras N años los trabajos publicados sean accesibles gratuitamente en formato digital. Lo usual es que N valga 5 o 10 años. Aunque no todo el mundo está de acuerdo con esta idea, el proyecto DML se moverá en esta dirección, salvo excepciones en casos específicos. Ya existe una recomendación del Comité Ejecutivo de IMU [4] para que las editoriales garanticen la libre disposición de la versión electrónica de los artículos *tras un período fijado de tiempo* desde su publicación. Para los libros, estaríamos hablando de un intervalo mucho más amplio, tal vez de 20 años [10]. La otra posibilidad, que se ha excluido en principio, sería crear unos acuerdos complejos de distribución y pago de derechos, pero probablemente ninguna organización pudiese hacer este esfuerzo a largo plazo.

DISPONIBILIDAD DEL MATERIAL No sólo hay que negociar permisos para digitalizar, sino para distribuir el resultado final y llevar un registro de todos los proyectos locales, en marcha o en preparación. ¿Cómo mantener la colección resultante? ¿Dónde archivarla? El problema a resolver consiste en [1]:

- distribuir muchas copias a diferentes sitios;
- suponer que al menos una copia sobrevivirá en el futuro;
- usarla para reproducir más copias.

En otras palabras:

- ¿En cuántos servidores (o *espejos*, es decir copias exactas de un servidor central) debe guardarse el material?
- ¿Será el formato de los ficheros compatible con futuros programas y sistemas operativos?
- ¿Cada cuánto tiempo deben hacerse copias nuevas? ¿Resistirá el soporte (magnético u óptico) el paso de los años?

En opinión de varios promotores del proyecto, su éxito depende de que se cree un depósito de *material en bruto* (*raw material*, es decir, los datos bibliográficos, las imágenes escaneadas, los ficheros de texto, pero sin estructurar) que sea accesible sin restricciones, para que distintas instituciones, cada una con sus propios intereses particulares (bibliotecas, sociedades matemáticas o editoriales), puedan crear diferentes mecanismos para dar valor añadido a la colección (por ejemplo, procedimientos para organizar el material, indexarlo, distribuirlo...), y mejoras de interés científico para acceder al material (posibilidad de referencias cruzadas y de búsqueda). Es este valor añadido el que tendría un precio, mediante el cobro de derechos de suscripción a determinados servicios.

En principio las editoriales asegurarán el archivo mientras sea económicamente viable, pero son las bibliotecas los lugares lógicos para archivar material y lo harán a largo plazo: *las bibliotecas son los guardianes naturales de los registros académicos, y han archivado tradicionalmente la literatura de investigación. Se trata de extender esta tradición de conservación, cuidado, archivo y acceso al entorno electrónico* [15].

FINANCIACIÓN Estamos hablando de aproximadamente 50 millones de páginas contenidas en libros, revistas y otros tipos de publicaciones. El coste estimado estrictamente para escaneo y reconocimiento óptico de caracteres (OCR) es de 1 euro por página, pero sumando gastos de personal y administración es probable que ascienda en total a unos 2 euros por página, así que el proyecto ronda los 100 millones de euros. Se calcula que un 20% del presupuesto estaría destinado a aspectos organizativos y administrativos.

En la primera fase el equipo de diseño tendrá que ser un consorcio internacional en el que participen todas las partes afectadas, y la financiación correrá por cuenta de diversas organizaciones de apoyo a la investigación. Por ejemplo, en EEUU estaba previsto cuadruplicar el presupuesto de la NSF para

Matemáticas, que pasaría de los 120 millones de dólares actuales a 500 millones cada año a partir de 2007; sin embargo las noticias más recientes son menos optimistas. En Europa, la EMS ha presentado una EoI (Expresión de Interés [9]) para el 6 programa marco de la UE, en el programa de apoyo a las infraestructuras de investigación.

La segunda fase, de digitalización, será un esfuerzo colectivo internacional. Muchos países disponen ya de fondos para digitalizar colecciones, y la idea es dividir el trabajo en varias tareas, cada una para una organización o país. Ya que el coste de digitalización es sólo por una vez, podrá ser financiado con las ayudas a la investigación de cada gobierno.

Tercero, el mantenimiento de una biblioteca electrónica global será responsabilidad de cada propietario o distribuidor y tendrá un coste, de archivo y acceso, que deberá ser cubierto por los usuarios o por las ayudas a la investigación de cada país. El aspecto más difícil del proyecto es estimar los costes y necesidades a largo plazo (por ejemplo transición a tecnologías y formatos futuros desconocidos), pero probablemente los gastos anuales de mantenimiento de una biblioteca digital serán un 15% del coste inicial.

Finalmente, el coste de organización del proyecto a largo plazo podría ser pagado por los miembros de DML a través de una cuota anual.

EDITORIALES En todo caso, se trata de diseñar un modelo económico que sea atractivo para las editoriales (no sólo grandes casas comerciales sino también sociedades matemáticas y universidades). Algunas ya están digitalizando sus colecciones y preferirán seguramente albergarlas en sus propios servidores y reservarse los derechos de autor, pero colaborarán si perciben algún beneficio o el proyecto se ajusta a sus objetivos globales. Otras muchas editoriales no pueden asumir el enorme gasto que supone desarrollar herramientas por sí mismas.

Los acuerdos deben proteger tanto los intereses de las editoriales como los objetivos de DML, y deben significar ventajas a cambio de ciertas obligaciones. Entre las primeras cabe citar: apoyo financiero para digitalizar; disponer de estándares mínimos que permitan la compatibilidad pero que sean suficientemente flexibles para adecuarlos a sus procedimientos actuales de publicación y de archivo; tener acceso a fondos nacionales e internacionales; aumento del tráfico en los servidores, pues los usuarios de DML serán dirigidos a sus portales en la web. Las editoriales retendrán su material y no se les pedirá transferir el *copyright*, salvo compartir la licencia para el formato electrónico con DML. A cambio tendrán que establecer secciones DML y albergar material en sus servidores; dar acceso y permitir enlaces de artículos y referencias a las bases de datos y a otros servidores; comprometerse a actualizar los formatos para asegurar la continuidad del sistema, y a que todo contenido digital nuevo se incorpore de forma rutinaria a DML.

4. PROYECTOS EN CURSO: EXPERIENCIAS

Existen proyectos ya en marcha en diversos países (tanto en América como en Europa), pero se trata de iniciativas a una escala mucho menor e iniciadas de manera descoordinada. Aunque existen contactos entre todos ellos, estos proyectos *no están basados en estándares comunes ni en principios generales, y probablemente utilizan formatos incompatibles* [21]. Sin ánimo de ser exhaustivos, podemos citar los siguientes.

NUMDAM [2, 13] En Francia, el programa NUMDAM (*Numérisation de Documents Anciens Mathématiques* <http://www-mathdoc.ujf-grenoble.fr/NUMDAM>) está financiado por el CNRS y gestionado por CelluleMathDoc (Grenoble), que es un grupo de coordinación documental formado por 3 matemáticos (a tiempo parcial), 2 ingenieros de computación, 2 documentalistas y un administrativo. Este proyecto comenzó en 2002 a digitalizar 220.000 páginas de cinco revistas matemáticas francesas de interés general, con unos 8.000 artículos aproximadamente. En este momento están completamente accesibles en línea los *Annales de Fourier* (1949–1996) y las *Journées Equations aux dérivées partielles* (1974–2000). Otras revistas previstas son las *Publications* del IHES, el *Bulletin* de la Soc. Mat. de Francia, o los *Annales* de la École Normale Sup., cada una con diferentes *moving walls*.

Puesto que este modelo podría ser aplicable en España, explicamos algunos detalles. Los objetivos de NUMDAM son proporcionar un acceso fácil a la herencia cultural matemática francesa, y combinar colecciones antiguas y publicaciones actuales. El acceso es gratuito (con el compromiso de un uso responsable), con enlaces a bases de datos, y el mantenimiento del archivo se hace en colaboración con algunas bibliotecas universitarias. Los métodos de digitalización usan estándares técnicos de alta calidad, y se garantiza la preservación de los derechos de autor mediante negociación con los responsables y editoriales de las revistas, a los que se pide autorización expresa, y se da un derecho compartido del *copyright* de los ficheros a cambio de la delegación del *copyright* del artículo original.

ERAM [27] En Alemania, el ERAM (*Electronic Research Archive for Mathematics*) constituye una biblioteca digital de importantes documentos matemáticos de los siglos XIX y XX. Incluye por ejemplo el archivo digital de la Univ. de Gottingen, con colecciones completas de prestigiosas revistas como *Math. Annalen* o *Math. Zeitschrift*, y los libros de la serie *Grundlehren der Mathematik*. Empezó en 1997, y lleva escaneadas 500.000 páginas, que esperan aumentar a 1.6 millones durante el año 2003. Forman parte del proyecto la EMS, *Zentralblatt* y editoriales alemanas. Las imágenes escaneadas están acompañadas de metadatos, facilidades de acceso, y pueden ser descargadas en formato PDF y conectadas a la base de datos ZMATH.

Contiene además la colección (desde 1868 hasta 1908) del *Jahrbuch über die Fortschritte der Mathematik* (JFM) y más de 12.000 enlaces a reproducciones facsímiles. JFM fué la primera revista de recensiones en Matemáticas y se publicó hasta 1942 (200.000 recensiones en total). Es accesible gratuitamente en el servidor EMIS de la EMS (<http://www.emis.de/MATHJFM/JFM>). Seis volúmenes (más de 6.000 páginas) han sido pasados a AMS-TEX por la empresa Balkan Data, Lefkowitz and Co. por encargo de la biblioteca de Gottingen (SUB).

Además SUB participa en otros muchos proyectos [1].

EUCLID [19] El proyecto EUCLID es una iniciativa conjunta de la Univ. de Cornell (Ithaca, New York, EEUU) y la Duke Univ. (Durham, North Carolina, EEUU), financiada por la fundación A.W. Mellon, en cooperación con editores independientes de revistas que están desarrollando el acceso digital. Estaba enfocada inicialmente a la publicación electrónica, pero trata de completar el proyecto JSTOR en cuanto a almacenamiento de números atrasados en formato digital. Este proyecto JSTOR (<http://www.jstor.org/>) negocia con editoriales para digitalizar colecciones completas de revistas, pero actualmente se centra más en otras disciplinas que en Matemáticas.

El primer proyecto de CUL (biblioteca de la Univ. de Cornell) fué la colección digital de libros de Matemáticas DMBC, que contiene en formato digital alrededor de 500 monografías anteriores a 1914 (puede accederse gratis en cdl.library.cornell.edu/math.html).

El proyecto DML en U.S.A. está coordinado por CUL y financiado por la NSF (National Science Foundation). El objetivo de nuevo es *establecer una colección completa, internacional y distribuida de información digital y de conocimiento publicado en Matemáticas* [5].

EMANI [12] Surgido a mediados de 2001, EMANI es un consorcio en el que cooperan bibliotecas (Cornell en EEUU, Gottingen en Alemania y Tsinghua en Pekín), y editores como Springer-Verlag y Birkhäuser. También participa ELibM, que es una biblioteca electrónica dependiente de la EMS en la que participan varias revistas y editores que ofrecen de manera gratuita contenidos electrónicos a través de servidores como EMIS (<http://www.emis.de>).

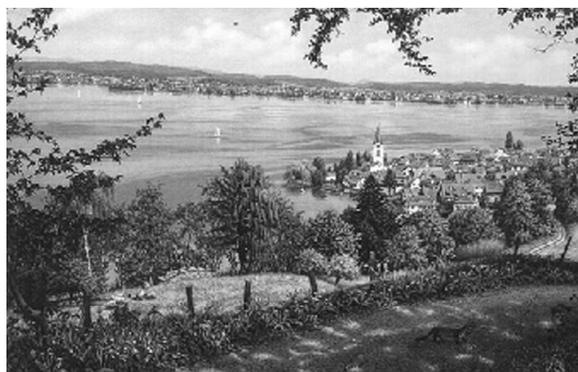
Se supone que el contenido de EMANI se almacenará en un depósito, y que se crearán copias en todas las bibliotecas de referencia por motivos de seguridad, intercambiándose periódicamente versiones actualizadas.

5. LA PARTICIPACIÓN EUROPEA

Está claro que Europa no puede quedar al margen de un proyecto así, dada la importancia de su legado matemático histórico, de la cantidad de revistas y libros que publica, y del tamaño global de sus bibliotecas. En una primera

aproximación, la contribución europea al proyecto consistiría en digitalizar 8 millones de páginas, con un costo estimado de 16 millones de euros.

La Sociedad Matemática Europea (EMS), en la que están representadas alrededor de 50 sociedades matemáticas, entre ellas la RSME, debe tener un papel destacado. De hecho, la EMS ha presentado una *expresión de interés (EoI)* [9] a la Comisión Europea para conseguir financiación dentro del 6 programa marco (6fp), y está preparando el correspondiente proyecto con el título de *DML-EU: A Digital Mathematics Library*.



La EMS preparó el proyecto DML-EU en Berlingen (Suiza)

La RSME ha participado en las reuniones preparatorias, y está diseñando en paralelo un plan de digitalización, en colaboración con la SCM (*Societat Catalana de Matematiques*) y la SPM (*Sociedade Portuguesa de Matemática*), en el que han mostrado su voluntad de participar las principales revistas de investigación españolas y portuguesas. Se ha empezado ya a acumular información sobre las revistas matemáticas del área, en términos de número de páginas, fechas de publicación y editoriales.

Debe tenerse en cuenta que un proyecto europeo no cubre los gastos del trabajo de digitalización en sí, sino que financia las iniciativas destinadas a: coordinar, facilitar y animar programas nacionales de digitalización; establecer modelos financieros fiables y viables que hagan atractivo a los propietarios de los contenidos digitalizar sus libros y revistas matemáticas, y distribuirlos a través de DML a un costo aceptable o gratis. El dinero europeo permite explorar caminos para digitalizar y hacer accesible el corpus bibliográfico, establecer herramientas técnicas comunes e integrar en una red europea de actividades los proyectos de digitalización existentes o en fase de proyecto. Recíprocamente será más fácil conseguir financiación nacional si hay un apoyo europeo visible, y si la comunidad matemática y las bibliotecas apoyan el proyecto.

CINDOC Cualquier centro que desee participar en el proyecto debe cumplir unos requisitos mínimos [23]. El primero, debe tener experiencia en actividades de digitalización o en revistas electrónicas, y de colaboración con organizaciones externas y participación en proyectos; experiencia en aspectos como OCR, metadatos y estándares; y una estructura sostenible para servir de depósito

y plataforma de distribución de documentos. Deberá disponer de unidades de digitalización modernas, como cámaras de alta resolución y escáners, OCR, y técnicas de manipulación de materiales raros o frágiles, y capacidad para desarrollar sistemas de producción para gestionar, preservar y distribuir a los usuarios finales el material digital y los metadatos asociados.

Para diseñar un plan de digitalización debe combinarse una buena infraestructura de base junto con experiencia en proyectos similares, en publicaciones matemáticas, y en el área de recuperación de información (por ejemplo catálogos con búsqueda y acceso *on-line* integrados, bases de datos, revistas electrónicas); hay que contar con expertos en T_EX, XML y otros formatos; y es fundamental tener el respaldo de sociedades matemáticas. Hay que contar con un lugar en donde pueda hacerse no sólo la digitalización de las revistas editadas aquí, sino que sirva de servidor espejo para la zona.

La RSME y la SCM consideran que el CINDOC (Centro de Información y Documentación del CSIC, <http://www.cindoc.csic.es/>) cumple estos requisitos, y han entablado negociaciones con él, así como con otros organismos estatales y autonómicos para diseñar un plan de digitalización y conseguir financiación adicional.

6. RESUMEN: CONCLUSIONES

De acuerdo con las recomendaciones de la UNESCO [16], el archivo digital de publicaciones electrónicas es esencial para que los resultados científicos no se pierdan para la posteridad. Distintas resoluciones del Consejo de la Unión Europea [6] han insistido en la necesidad de digitalizar los contenidos culturales con el fin de conservar el patrimonio europeo.

Hay un sentimiento general de que la literatura matemática debería ser accesible libremente y gratis. La necesidad de disponibilidad a largo plazo es típica de las Matemáticas. La documentación matemática es especialmente duradera, razón por la que debe ser conservada y hecha accesible fácilmente.

Es urgente poner orden en la cantidad cada vez mayor de proyectos locales de retro-digitalización y archivo de publicaciones electrónicas, sin coordinación entre ellos, basados en estándares técnicos diferentes, en formatos diferentes, y con interfaces diferentes,

La creación de una biblioteca digital virtual con toda la literatura académica será parte fundamental de la infraestructura mundial para las Matemáticas y áreas afines [25]. El proyecto requiere una preparación muy cuidadosa, la participación de los mejores recursos y una financiación suficiente [11]. El resultado será una colección digital representando una cantidad significativa de nuestra herencia cultural científica, con un sistema de distribución a través de bibliotecas universitarias y acceso de los usuarios mediante procedimientos establecidos.

El proyecto es del máximo interés para los países menos desarrollados, donde el acceso a la literatura matemática es actualmente difícil o inexistente.

Dada la importancia de las Matemáticas en el desarrollo científico y tecnológico, el impacto del proyecto DML se notará también en las otras Ciencias.

En palabras de Ph. Tondeur [25]: *Si tenemos éxito con este proyecto, su impacto en las Matemáticas, en las ciencias y en la educación puede que sea el más relevante desde la invención de las revistas científicas como substituto de la correspondencia privada entre científicos individuales.*

REFERENCIAS

- [1] H. J. BECKER, *Archiving, Library Gottingen SUB*, EMS Digitization Meeting, Berlingen (Suiza), Enero 2003.
- [2] P. BÉRARD, *NUMDAM, Digitization of ancient mathematical documents*, Joint Math. Digitization Meeting, San Diego, California, Enero 2002.
- [3] TH. BOUCHE, U. REHMANN *et alt.*, *Intermediate report of the DML working group on technical standards*, Nov. 4, 2002.
<http://www.mathematik.uni-bielefeld.de/rehmann/DML/>
- [4] CEIC (IMU Committee on Electronic Information and Communication), *Best current practices: Recommendations on Electronic Information Communication (2002)*. Aprobado por el Comité Ejecutivo de IMU, sesión 69, 13 Abril 2002, París.
<http://www.mathunion.org>
- [5] <http://www.library.cornell.edu/dmlib/>
- [6] Resolución 2002/C 162/02 del Consejo de la Unión Europea, Conservar la memoria del mañana: Conservar los contenidos digitales para las generaciones futuras, 25/6/2002
http://europa.eu.int/eur-lex/es/archive/2002/c_16220020706es.html
- [7] J. CRAWLEY *et alt.* *DML working group on economic models*, SIAM-CMS-AMS-CUP-Springer, Enero 2003.
- [8] A. DE ROBBIO, *Digital Math Library DML: un progetto mondiale per la matematica*, Technical Report, Biblioteca del Seminario Matematico, Universita degli Studi di Padova (2003)
<http://eprints.rclis.org/archive/00000099/>
- [9] Digital Mathematics Library, Expression of Interest to the European Union submitted by the European Mathematical Society <http://eoi.cordis.lu/>
<http://www.maths.leeds.ac.uk/pure/staff/salinger/EMS/DML.pdf>
- [10] J. EWING, *Twenty centuries of mathematics: Digitizing and disseminating the past mathematical literature*, Marzo 2002
http://www.ams.org/ewing/Twenty_centuries.pdf
- [11] M. GRETSCHEL, Joint Mathematics Meetings in San Diego on Jan. 7, 2002.
<http://www.library.cornell.edu/dmlib/SDsummaries.html#SDsumm>

- [12] G. GRIEPKE, B. WEGNER, *The Electronic Mathematics Archive Initiative (EMANI)*, Springer-Verlag & TU Berlin, Joint Math. Digitization Meeting, San Diego, California, Enero 2002.
- [13] L. GUILLOPÉ, *The NUMDAM program, CelluleMathDoc-Grenoble*, EMS Digitization meeting, Berlingen (Suiza), Enero 2003.
- [14] J. HEINZE, 2002
<http://www.online-information.co.uk/online/proceedings/wegner.pdf>.
- [15] IFLA (International Federation of Library Associations and Institutions); IPA (International Publishers Association), *Preserving the memory of the world in perpetuity: a joint statement on the archiving and preserving of digital information*, Paris 2002.
www.ifla.org
- [16] ICSU-UNESCO International Conference on Electronic Publishing in Science, Paris 2001
http://www.unesco.org/science/publication/electronic_publishing_2001/proceedings.htm
Ver también
http://www.unesco.org/science/publication/eng_pub/elecpub.htm
- [17] ISO Archiving Standards, 2001
<http://ssdoo.gsfc.nasa.gov/nost/isoas/overview.html>.
- [18] M. JONES, C. WEBB, *Electronic Publications and the Survival of Information, National Library of Australia*
<http://www.nla.gov.au/nla/staffpaper/mjcwala.html>
- [19] Z. KOLTAY, *Project Euclid, Cornell University*, Joint Math. Digitization Meeting, San Diego, California, Enero 2002.
- [20] G. KUPERBERG, Scholarly mathematical communication at a crossroads, *Nieuw Arch. Wisk* 5/3, nr. 3 (2002), 262–264.
- [21] L. Lemaire, *A report of EMS for OECD, Access to publicly funded research in mathematics. A Digital Mathematics Library*. Univ. Libre de Bruxelles, Enero 2003.
- [22] MARKVORSEN, *The math of math literature*, EMS Digitization Meeting, Berlingen (Suiza), Enero 2003.
- [23] D. PRICE, *Digital Mathematics Library, Potential involvement of Oxford Library Services, UK*, Systems & Electronic Resources Service, Oxford, Enero 2003.
- [24] M. SANDFAER, L. FUGL, *Ideas from Technical University of Denmark, Center for Knowledge Technology/InfoLab and Department of Mathematics*, EMS Digitization Meeting, Berlingen (Suiza), Enero 2003. (DTV), Jan. 2003.
- [25] PH. TONDEUR, *Digital Mathematical Library*, Joint Math. Digitization Meeting, San Diego, California, Enero 2002.

- [26] Por ejemplo
<http://any2djvu.djvuzone.org/>
transforma cualquier formato a djvu;
<http://www.mathematik.uni-osnabrueck.de/cgi-bin/MM3.1cgi>
es un prototipo de servidor MathNet para metadatos.
- [27] B. WEGNER, *ERAM*, Joint Math. Digitization Meeting, San Diego, California,
Enero 2002.
<http://www.library.cornell.edu/dmlib/SDsummaries.html#SDsumm>

Enrique Macias-Virgós
Comisión de Cooperación Internacional de la RSME
Departamento de Geometría y Topología
Facultad de Matemáticas
Universidad de Santiago de Compostela
Campus Sur
15786 - Santiago de Compostela
Correo-electrónico: xtquique@usc.es