

The Digital Mathematics Library

Not long ago, Keith Dennis, a mathematician at Cornell University, walked into the departmental photocopying room and saw a bunch of old journals with articles tabbed for photocopying. He told the secretary assigned to make the copies that the journals are available electronically through the JSTOR journal storage website. She was delighted not to have to spend time standing in front of the copy machine. For his part, Dennis was puzzled that one of his colleagues evidently did not realize how easily one can get the material on the Web. "You would hope mathematicians would have some idea of where their literature is," he said. "But that's simply not true."

Dennis told this anecdote at a meeting of the Digital Mathematics Library (DML) Planning Group that took place in May this year in Göttingen, Germany. Despite the committee's work, which was supported by a one-year grant from the National Science Foundation (NSF), and despite progress in getting older paper literature online, the mathematical community remains largely unaware of what is available through the Web. Back issues of such major journals as *Inventiones Mathematicae*, *Mathematische Annalen*, and *Publications Mathématiques de l'IHÉS* were recently made available for free online, and there is more to come. This work is mostly being carried out by a few disparate projects that operate independently. The challenge now facing the world mathematical community is not only to secure funding for such projects but also to come up with ways to coordinate them to prevent duplication of effort, encourage adherence to standards, and ensure that the material is widely accessible.

Pursuing a Dream

The grand vision of the DML is to have all of the mathematical literature online and available through a central source to anyone who has a computer and an Internet connection. Much of the current literature is "born-digital", that is, created electronically and available online from the time of publication. Although high subscription fees can hinder access, that material at least exists in electronic form. But the vast majority of mathematics books and journals are still available only on paper. "Retrodigitization" is the name for the process of creating electronic copies of paper-only works. The initial goal of the DML is the retrodigitization of all of the past mathematics literature.

This goal was strongly pushed by Philippe Tondeur, now retired from the University of Illinois at Urbana-Champaign, who served as director of the NSF's Division of Mathematical Sciences from 1999 until 2002. He used the bully pulpit of his NSF position to proselytize widely for the DML. After discussing the idea with science funding officials in several countries, including officials at the European Commission, Tondeur felt that there was sufficient worldwide support to undertake a program of retrodigitizing all of mathematics. The NSF then provided a one-year grant to Cornell University Library to support meetings of what came to be called the DML Planning Group (the co-principal investigators on the grant were Keith Dennis and Jean Poland, Cornell's associate university librarian for engineering, mathematics, and physical sciences). The group included a steering committee as well as six working groups

and two liaisons to the International Mathematical Union (IMU) (see sidebar). Three meetings were held, the most recent on May 21–22, 2003, at the Niedersächsische Staats- und Universitätsbibliothek in Göttingen.

To explore the potential and challenges of the DML vision, Tondeur asked AMS executive director John H. Ewing to write a “concept paper”, and this paper came to be widely circulated (“Twenty centuries of mathematics: Digitizing and disseminating the past mathematical literature”, *Notices*, August 2002, pages 771–7). Among the hurdles identified in the paper were negotiating permissions for copyrighted works, making editorial decisions about which books and papers should be included, choosing storage formats, and archiving over the long term. Ewing argued that, despite those difficulties, the grand vision of the DML is feasible: with today’s technology, it is actually a tractable task to put all 50 million pages of the past mathematical literature online.

One reason the DML idea has gained momentum is that, as Ewing’s paper points out, mathematics is an ideal discipline for massive retrodigitization. In a sense, mathematics is indistinguishable from its literature. Unlike researchers in many other disciplines, especially in the sciences and engineering, mathematicians rely heavily on past literature while working at the frontiers of research. Having that literature available electronically would have a large impact on current research in mathematics.

On the other hand, compared to other disciplines, mathematics presents special challenges for retrodigitization. For one thing, the mathematics literature is extremely diffuse. Dennis, who served as executive editor for *Mathematical Reviews* (MR) from 1995 until 1998, recalled an early conversation with the developers of JSTOR, which has retrodigitized journals in mathematics and other areas. They asked Dennis how many mathematics journals they should put online. Jaws dropped when he replied, “Let’s start with five hundred.” He was not joking: About six hundred mathematics journals are treated cover-to-cover by MR, and mathematical items are chosen for review from hundreds of other journals not exclusively devoted to mathematics. In some other disciplines, by contrast, it would suffice to retrodigitize as few as a dozen of the most important journals. What is more, mathematics journals are published by commercial publishers, university presses, professional societies, mathematics departments, even ad hoc groups of mathematicians. It would be an enormous legal task to negotiate copyright agreements with such a diverse and geographically dispersed group.

Digital Mathematical Library Planning Group

Steering Committee

Hans Becker
Niedersächsische Staats- und Universitätsbibliothek
in Göttingen

Pierre Bérard
Université Joseph Fourier, Grenoble

Keith Dennis
Cornell University

Jean Poland
Cornell University Library

Bernd Wegner
Technische Universität Berlin

IMU Liaison Committee

Rolf Jeltsch
Eidgenössische Technische Hochschule Zürich

David Mumford
Brown University

Working Groups

Content

Keith Dennis
Steve Rockey
Cornell University Library

Bernd Wegner

Technical Standards

Thierry Bouche
Université Joseph Fourier, Grenoble

Ulf Rehmann
Universität Bielefeld

Metadata

Tim Cole
University of Illinois, Urbana-Champaign

Heike Neuroth
Niedersächsische Staats- und Universitätsbibliothek
in Göttingen

Robbie Robson
Eduworks Corporation

Rights and Licenses

Pierre Bérard
David Tranah
Cambridge University Press

Archiving

Hans Becker
Kizer Walker
Cornell University Library

Economic Model

Jonathan Borwein
Simon Fraser University

James Crowley
Society for Industrial and Applied Mathematics

John Ewing
American Mathematical Society

Arnoud de Kemp
Springer-Verlag

David Tranah
Cambridge University Press

Retrodigitized Mathematics Journals

Below is a listing of retrodigitized mathematics journals available on the Web. Not included here are websites providing "born digital" material.

Biblioteka Wirtualna Matematyki

<http://matwbn.icm.edu.pl/>
Fundamenta Mathematicae (1920–1993)
Studia Mathematica (1929–1964)
Prace matematyczno-fizyczne (1888–1952)

Departamento de Ingeniería Matemática

Universidad de Chile
<http://www.dim.uchile.cl/revmat.html>
Revista de Matemáticas Aplicadas (1994–2002)
(complete retrodigitization under way)

DIEPER

DIGitised European PERiodicals
<http://dieper.aib.uni-linz.ac.at>
Monatshefte für Mathematik und Physik (1890–1918)

EMIS

European Mathematical Information Service
<http://www.emis.de/>
Jahrbuch über die Fortschritte der Mathematik (1868–1931)

Gallica

Bibliothèque nationale de France
<http://gallica.bnf.fr>
Journal de Mathématiques Pures et Appliquées (1836–1880)
(access available through <http://www-mathdoc.ujf-grenoble.fr/jmpa>)
Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences (1835–1930)

JSTOR

Journal STORage
<http://www.jstor.org>
(paid subscription required)
American Journal of Mathematics (1878–1995)
American Mathematical Monthly (1894–1997)
Annals of Applied Probability (1991–1997)
Annals of Mathematical Statistics (1930–1972)
Annals of Mathematics (1884–1997)
Annals of Probability (1973–1997)
Annals of Statistics (1973–1997)
Applied Statistics (1952–1998)
Biometrika (1901–1997)
Bulletin of Symbolic Logic (1995–2002)
College Mathematics Journal (1984–1997)
Econometrica (1933–1998)
Journal of the American Statistical Association (1922–1997)
Journal of the AMS (1988–1997)
Journal of the Royal Statistical Society. Series A (Statistics in Society) (1988–1998)
Journal of the Royal Statistical Society. Series B (Statistical Methodology) (1998)
Journal of Symbolic Logic (1936–1998)
Mathematics Magazine (1947–1997)
Mathematics of Computation (1960–1997)
Philosophical Transactions: Mathematical, Physical and Engineering Sciences (1665–1997)
Proceedings: Mathematical, Physical, and Engineering Sciences (1800–1997)
Proceedings of the AMS (1950–1997)
SIAM Journal on Applied Mathematics (1966–1997)
SIAM Journal on Numerical Analysis (1966–1997)
SIAM Review (1959–1997)
Statistical Science (1986–1997)
The Statistician (1962–1998)
Transactions of the AMS (1900–1997)

NUMDAM

NUMérisation de Documents Anciens Mathématiques
<http://www.numdam.org>
Annales de l'Institut Fourier (1949–1997)
Annales Scientifiques de l'École Normale Supérieure (1864–1997)
(to be added fall 2003)
Bulletin de la Société Mathématique de France (1872–1992)
Journées Équations aux dérivées partielles (1974–2000)
Mémoires de la Société Mathématique de France (1964–1992)
Publications Mathématiques de l'IHÉS (1959–1997)

Project Euclid

<http://ProjectEuclid.org/mmj>
Michigan Mathematical Journal (1952–present)

SciELO

Scientific Electronic Library Online
<http://www.scielo.cl/scielo.php>
Proyecciones—Revista de Matemática (2000–2002)

WDMG-Göttingen

Göttinger Digitalisierungszentrum
<http://www.sub.uni-goettingen.de/gdz>
Abhandlungen der Gesellschaft der Wissenschaften in Göttingen, Mathematisch-Physikalische Klasse (1900–1939)
Abhandlungen der Königlichen Gesellschaft der Wissenschaften in Göttingen (1843–1892)
Acta Facultatis Rerum Naturalium Universitatis Comenianae (1956–1975)
Acta mathematica Universitatis Comenianae (1980–1980)
Aequationes mathematicae (1968–1997)
Archivum mathematicum (1965–1991)
Beiträge zur Algebra und Geometrie (1971–1992)
Casopis pro pestování matematiky (1951–1990)
Casopis pro pestování matematiky a fyziky (1872–1950)
Commentarii mathematici Helvetici (1929–1937/38)
Commentationes mathematicae Universitatis Carolinae (1960–1990)
Geometric and functional analysis (1991–1996)
Inventiones mathematicae (1966–1996)
Matematicki vesnik (1964–1995)
Mathematica Bohemica (1991–1994)
Mathematica Scandinavica (1953–1957)
Mathematische Annalen (1869–1996)
Mathematische Zeitschrift (1918–1996)
Mémoires de l'Académie Royale des Sciences, des Lettres et des Beaux-Arts de Belgique (1777–1788; 1847–1897)
Metrika (1958–1971)
Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse (1895–1933)
Nachrichten von der Königl. Gesellschaft der Wissenschaften und der Georg-Augusts-Universität zu Göttingen (1865–1893)
Nouveaux mémoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles (1820–1845)
Numerische Mathematik (1959–1982)
Revista colombiana de matemáticas (1967–1993)
Seminaire de Théorie des Nombres de Bordeaux (1971–1988)
Vesnik Društva Matematicara i Fizicara Narodne Republike Srbije (1949–1963)
Zentralblatt für Mathematik und ihre Grenzgebiete (1931–1978)

Examples of Retrodigitization Projects

Much of the process of turning paper into electronic files can be automated. Generally, the material is not retyped. Rather, each page is scanned to create an electronic “picture” of the page. Often optical character recognition software is used to produce a text file from the scanned image so that the actual text of the material is searchable electronically. Sometimes the search feature highlights the sought-after word rather than just displaying the page on which the word appears. Users typically retrieve the material in the form of PDF files; other formats such as PostScript or DjVu¹ are sometimes available. Usually the bibliographic data must be typed to ensure standardization and accuracy.

So far retrodigitization projects in mathematics have tended to be fairly small and regionally based. Three examples are JSTOR in the U.S., NUMDAM in France, and WDML-Göttingen (World Digital Mathematical Library) in Germany. JSTOR, which began in 1997, currently offers seventeen journals in the mathematical sciences, most of them based in the U.S., including the *Annals of Mathematics*, *Journal of the AMS*, *Transactions of the AMS*, and the *American Mathematical Monthly*; there are also thirteen journals in statistics and probability. The text of the articles is fully searchable, and one can download them in TIFF, PDF, and PostScript formats. Like most retrodigitization projects, JSTOR operates with a “moving wall”: Only material that has been out more than a certain number of years can appear on JSTOR. The number of years varies from journal to journal but is generally three to five years. Publishers have demanded such policies in order to protect journal subscription revenues.

NUMDAM (NUMérisation de Documents Anciens Mathématiques), based at the Université Joseph Fourier in Grenoble and supported by the Centre National de la Recherche Scientifique (CNRS), came online in December 2002. It offers back issues of five French journals, and there are plans to add more. One can find on NUMDAM, for example, the complete text of the landmark work *Éléments de Géométrie Algébrique*, by Alexandre Grothendieck and Jean Dieudonné, which appeared in *Publications Mathématiques de l’IHÉS* in the 1960s. Articles are available in PDF and DjVu formats, and the text is fully searchable. NUMDAM has a feature similar to the “author identification” feature of MathSciNet: there is a complete alphabetical listing of all authors, organized to account for variations in the spelling and presentation of the authors’ names. To the extent possible, NUMDAM has added links

¹The DjVu format, though less well known than PostScript and PDF, is in some ways superior. More information about it may be found at <http://www.djvuzone.org>. The site includes convenient tools for converting from other formats to DjVu.

from articles to reviews in MathSciNet (the online version of MR) and to reviews in Zentralblatt MATH (the online version of *Zentralblatt für Mathematik und ihre Grenzgebiete*). NUMDAM also provides links from references within articles to reviews in MR and *Zentralblatt*.

The other European example, WDML-Göttingen, is part of a retrodigitization center based at the Niedersächsische Staats- und Universitätsbibliothek in Göttingen. This center, the Göttinger Digitalisierungszentrum (GDZ), currently offers access to 4,100 volumes and 1.5 million pages (900,000 of them mathematics), including such culturally significant works as the Gutenberg Bible. The capabilities of WDML-Göttingen are more limited than those of JSTOR and NUMDAM: for one thing, optical character recognition has not been performed on the files, so the actual text of the articles is not searchable (though one can search bibliographic and structural metadata, and there are special tools to navigate within documents). Nevertheless, WDML-Göttingen has brought a large amount of significant material to the Web. It offers twenty-eight journals, including complete pre-1996 runs of such Springer-Verlag journals as *Inventiones Mathematicae*, *Mathematische Annalen*, and *Mathematische Zeitschrift*. In addition, there are almost four hundred monographs and about twenty multivolume works, including the collected works of Carl Friedrich Gauss, Felix Klein, and David Hilbert, and both the 1898 and the 1939 editions of the *Encyklopädie der Mathematischen Wissenschaften mit Einschluss ihrer Anwendungen*. At the meeting of the DML Planning Group, hopes were high that *Crelle’s Journal (Journal für die Reine und Angewandte Mathematik)* would soon be added.

The appearance of the Springer journals on WDML-Göttingen is the result of a project called EMANI (Electronic Mathematical Archiving Network Initiative), which aims to make retrodigitized and born-digital materials available on the Web. Bernd Wegner of the Technische Universität in Berlin, who is editor of *Zentralblatt*, is a main mover behind EMANI. He helped to negotiate the terms with Springer, a commercial publisher in Germany. Similarly, French mathematicians associated with NUMDAM were successful in persuading some French publishers to allow retrodigitization of their journals. Local knowledge was essential in the success of these negotiations. Other commercial publishers, such as Elsevier, have begun their own in-house retrodigitization programs, access to which is not free of charge.

NUMDAM and WDML-Göttingen have received support from the science funding agencies of the French and German governments—the CNRS in France, and the Deutsche Forschungsgemeinschaft in Germany. In the U.S., despite the NSF’s support

Retrodigitized Mathematics Books

The Cellule MathDoc at the Université Joseph Fourier in Grenoble provides a centralized catalogue for digitized books in four locations: the Digital Math Books of the Cornell University Library, the mathematics books in the Gallica Collection of the Bibliothèque Nationale de France, the WDML-Göttingen project at the Göttinger Digitalisierungszentrum, and the University of Michigan Historical Mathematics Collection. See <http://math-doc.ujf-grenoble.fr/LiNuM/>.

Some monographs by Polish mathematicians are available on the Biblioteka Wirtualna Matematyki, <http://matwbn.icm.edu.pl/>.

for the DML Planning Group, there is essentially no federal funding for such projects. JSTOR got its start with funding through a private foundation, the Andrew W. Mellon Foundation, and, unlike NUMDAM and WDML-Göttingen, charges access fees. These fees run into the thousands of dollars, making JSTOR difficult for many institutions to afford. Nevertheless, JSTOR provides a successful model of a self-supporting, nonprofit organization that raises enough money through access fees to support ongoing maintenance and continued expansion of its database.

The economics of retrodigitization are somewhat surprising. Scanning in printed material is actually quite cheap. Many retrodigitization projects send material to low-wage countries to be scanned. But even the GDZ, which is located in the exceptionally high-wage country of Germany, can hire local workers for a reasonable cost to scan material on a per-page basis. In fact, the lion's share of the cost of retrodigitization is outside of the scanning step; without further processing, the actual content of the scanned works is largely inaccessible. As Ewing put it during the Göttingen meeting, "If you just have a bunch of images, you can't do anything."

During the meeting, representatives from the GDZ outlined the entire process whereby they convert paper materials into accessible electronic archives. Even without performing optical character recognition on the materials they scan, there is a substantial amount of work to be done in performing quality control and in collecting, organizing, and managing bibliographic data and information about the structure of the documents. Creating and maintaining long-term archives is another costly task. Because the GDZ is embedded in a large library, it is difficult to obtain precise estimates of per-page costs of the documents they retrodigitize. But when the figure of \$2 per page was mentioned during the Göttingen meeting, one

of the GDZ team members said that figure was in line with some early estimates they had done. The scanning might comprise just 10 percent of that amount.

Uniting the Literature

As helpful as resources like JSTOR, NUMDAM, and WDML-Göttingen are, mathematicians do not want to have to stop and remember whether a particular journal is on this or that server. What is needed is some kind of centralized access. One natural idea, which was discussed at the Göttingen meeting, is to add the necessary links to the two main bibliographic databases, MathSciNet and Zentralblatt MATH. In fact, this has already begun to happen. For example, for any paper that has been reviewed in MR and is available on JSTOR, MR has added links so that one can click directly from the review on MathSciNet to the paper on JSTOR (assuming one is at an institution subscribing to JSTOR). Many of the papers on JSTOR appeared before MR began in 1940, so those papers have no bibliographic records in MathSciNet. MR has begun the process of adding these records to MathSciNet, together with links to the papers in JSTOR. This process has been completed for all pre-1940 papers in *Transactions of the AMS* and is under way for *Annals of Mathematics* and for the *American Journal of Mathematics*. Similarly, links have been added from Zentralblatt MATH and from the *Jahrbuch* reviewing journal (which has been retrodigitized and is available online) to the materials available on WDML-Göttingen. At the Göttingen meeting, there was an enthusiastic consensus that such linking should be expanded as much as possible. During the meeting a small group of representatives from MR, *Zentralblatt*, and some retrodigitization projects agreed to confer on developing technical standards to facilitate this linking.

As the DML is now taking form, it consists of a collection of disparate projects working independently. Would it perhaps make sense to create a central body that would coordinate the entire DML retrodigitization program? Such an approach was described in Ewing's paper and discussed within the DML Planning Group, but it did not take root. Those running existing retrodigitization projects want to continue their work as they see fit rather than follow rules set by a larger authority. Similarly, the idea of raising money for the DML in a centralized way, by asking publishers to contribute 1 or 2 percent of their journal revenues, was discussed by the planning group and abandoned. It was assumed that publishers would simply raise subscription prices to cover the contribution. That the group would propose a plan that could lead to increased journal prices was anathema to the group members.

Nevertheless, the DML Planning Group, through its meetings, has already stimulated better coordination among the various retrodigitization projects. It has also produced a report containing contributions by each working group, thereby bringing together the thoughts and ideas of some of the members of the international mathematical community who are the most knowledgeable about retrodigitization. The report, submitted to the NSF in June, will likely prove useful in helping new projects get off the ground and in providing a starting point for standards. Although the planning group disbanded at the end of the Göttingen meeting, there was a clear consensus that a new body was needed to continue the discussion and coordination the group had begun. This task will now be taken up by the IMU Committee on Electronic Information and Communication (CEIC). The CEIC is considering organizing a meeting about the DML sometime within the next year.

Another coordinating body may emerge from a new retrodigitization program proposed in Europe. In April 2003 a group of European mathematicians submitted a proposal to the European Commission for funding for a DML in Europe, to be called DML-EU. Rolf Jeltsch of the Eidgenössische Technische Hochschule Zürich, who was an IMU liaison for the DML Planning Group, is the principal investigator on the proposal. The five-year project would involve more than forty groups in about two dozen nations across Europe. The idea is to stimulate in the individual countries new retrodigitization projects, so that each country takes responsibility for converting journals based in that country. The grant would fund only research and start-up efforts; the actual retrodigitizing would be financed by the individual countries. The proposal requests 7.9 million euros (about US\$9.2 million) from the European Commission; another 2.4 million euros is expected to be contributed by the individual countries. A preliminary meeting for the DML-EU was held May 22–23, 2003, immediately following the meeting of the DML Planning Group; in fact, many members of the group stayed for the DML-EU meeting.

The DML-EU proposal describes the formation of an entity provisionally called the “World Mathematical Library Club.” The word “club” is intended to suggest, as the proposal puts it, that the entity “does not itself control the digitization projects nor [sic] the contents.” Rather, the club would encourage worldwide coordination of retrodigitization projects, including those under the DML-EU. The club would have representatives from various groups having a stake in retrodigitization efforts: publishers, mathematical societies, libraries, retrodigitization projects, and so forth. In addition to providing a forum for communication among

these groups, the club would, according to the proposal, “function as the body to approve guidelines for the digitization, codes of conduct and so on.” The government of Finland has indicated willingness to host the club, provide it with legal status, and support a small office. Membership fees may be collected to provide financial support for the club.

Momentum Is Building

During the Göttingen meeting, IMU president John Ball of Oxford University said that the IMU views the DML as a “vital effort for the mathematical community.” He also noted that a good deal of momentum has now built up for the DML, “for mathematicians and for funding agencies,” and he urged that this momentum not be lost. Indeed, this is a critical moment in the development of the DML. The next steps taken by the international mathematical community may determine the future of this emerging resource, which could have a profound impact on mathematics research.

—*Allyn Jackson*