

# The WDML – vision and reality

The aim of the World Digital Mathematics Library WDML to make all mathematical publications electronically available. FIZ Karlsruhe contributes to this global enterprise by maintaining and offering the Electronic Library of Mathematics in EMIS (European Mathematical Information Service). This is an open access library with about 40 copies worldwide.

Bernd Wegner

## 1. The aim of the WDML

Mathematical publications are representing what may be considered as the current mathematical knowledge. The mathematics libraries are the laboratory for the research mathematician and people who want to apply the mathematical findings available at present. Investigations of new subjects and research on open problems in mathematics have to take into account all publications which are relevant for the topic under consideration. Good access to the existing mathematical literature is an essential requirement for an environment enabling good mathematical research and successful application of mathematical findings to solving problems in other sciences.

Confronted with the tremendous number of existing mathematical research articles a library has no chance

to keep printed records for all of them. Even libraries which are supposed to collect everything appearing in some part of mathematics have difficulties to acquire all relevant publications. An additional problem is given by the longevity of mathematical publications. The validity of the majority of findings has no expiration date. Statistics evaluating the citations in current publications show a surprisingly low decrease of older citations.

In principle, the possibility to post digital publications on the web shows a way how to get out of this dilemma. There is the chance to establish a comprehensive library in mathematics using electronic media. This idea is called the WDML in the following. In 2001 John Ewing published a study on the WDML (Twenty Centuries of Mathematics: Digitizing and disseminating the past mathematical literature. [www.ams.org/ewing/Twenty\\_centuries.pdf](http://www.ams.org/ewing/Twenty_centuries.pdf)). He addressed some first fundamental problems arising with the potential development of the WDML and he also mentioned an amount of 100 million USD as a rough estimate for the costs to be spent for the retrospective digitization of all mathematical publications available in printed form only.

The NSF took this report as the basis for funding a project at the library of the Cornell University for the years 2002 and 2003 with the aim to prepare a feasibility study on the WDML. The project involved partners from all over the world. The result has been published on the web: [www.library.cornell.edu/dmlib/](http://www.library.cornell.edu/dmlib/). The further monitoring of a potential implementation of the WDML was adopted by the IMU (International Mathematical Union). But this only had marginal impact on further steps undertaken by the owners of mathematical publications.

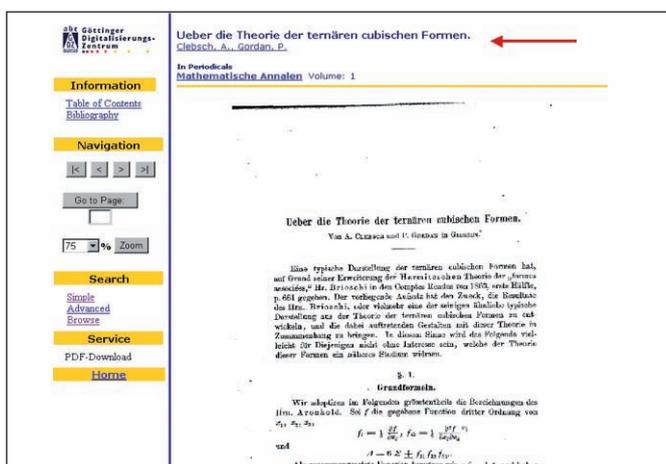


Fig. 12: Digitized article in ERAM. Low resolution choice for quick data transfer

The first patches of the WDML were results of digitization projects, where publishers, agencies, libraries or societies succeeded in getting some money for digitizing a part of the mathematical literature. The coordination between the projects was weak, if there was any at all. Later projects took advantage of some standards available from the NSF study. But even until now a main feature of the WDML, the integrated access, is not available, neither on the global scale, nor for any two of these projects. To establish this is delegated to additional projects. As an intermediate solution the links from Zentralblatt MATH and MathSciNet may serve as a replacement for this facility. A global solution will not be confronted with technical problems only. A hard obstruction for an integrated access is given by the different ways for subscribing to the distributed services. This ranges from open access to high priced controlled access.

## 2. What should be in the WDML?

A lot of time during the NSF project was spent for discussing governance problems. But this discussion was meaningless as long as it was not clear who should be governed, and the latter question is related to the question what the content of the WDML and the schedule for generating the content should be. All these questions became less important after the participants in the project had learned that NSF only would fund the feasibility study, but would not spend any money on the digitization itself.

The problem of identifying the desirable content remained of independent interest, because it is an essential part of the definition of the WDML. Several parameters had been developed for this purpose. One of them is time. How far should we go back in time? The reply to this has been anticipated by the Chinese. They are the most advanced community as refers to digitizing mathematical publications. For the Chinese mathematical publications everything is done already, including monographs and old Chinese mathematical documents. Another parameter is mathematics. When do we classify the topic of a publication as mathematics? There is no clear reply to this question. The mathematical reviewing services permanently are confronted with the same question, and they decided to accept a fuzzy borderline. Here the WDML has to cooperate with several WDLs. As a consequence retrospective digitization of multidisciplinary journals should cover whole volumes, non-mathematical papers included.

Another parameter will be the level of the publication. Should the WDML only store research publications? If so, what is the attitude referring to journals without a peer-reviewing system, preprints, rare manuscripts, and other grey literature? Should educational material and popularization of mathematics really be excluded? At present the reply is given by the decision of the funding organizations, and it changes from project to project.

## 3. Digitization projects

At present the retrospective digitization of publications available in printed form only will generate the core of the WDML. The production of digitally born publications in mathematics on a larger scale started in the middle of the nineties. Most of them can look back to a “printed only” period. Hence, as soon as that part will have been retrodigitized and interlinked with the digitally born period in a seamless way, the full range of the journal may be considered as part of the WDML.

The first initiative to digitize scientific journals at a larger scale was pursued by JSTOR. For the mathematics part they stored the most important Anglo-American journals in their digital archive. Bigger portions of mathematical publications were digitized by projects at Elsevier and Springer. They covered their own journals completely. At present Springer is extending these activities to their book production. Smaller publishers and societies possessing a relevant set of mathematical journals like SIAM und LMS caught up soon. They have in common that customers have to pay for the access.

Open access is typical for digital archives which developed their digital holdings with support from government resources or funding from foundations. In Germany the DFG gave the money for pursuing the projects ERAM and RusDML at SUB Göttingen. ERAM also enabled the structured capture of the content of the Jahrbuch über die Fortschritte der Mathematik. The format for the storage was the same as the one used at Zentralblatt MATH. Hence the data easily could be integrated later into the offer of Zentralblatt. RusDML was a Russian-German project with the aim to develop a digital archive of Russian publications in mathematics with a bilingual access structure.

Fig. 13: Entry page of RusDML

The digitization of the French mathematical journals was subject of the government funded project NUMDAM at UJF in Grenoble. In Europe during the last years several national projects lead to national DMLs like in Poland, Czech Republic, Spain, Serbia, Bulgaria, Portugal and Switzerland. On the world wide level the development of digital archives in China, Korea and the USA has to be mentioned. Several years ago KISTI had digitized all of the 16 mathematical journals published in Korea. As already mentioned above, there was a huge digitization project in China covering in addition to more than 50 mathematical journals also books and documents dealing with old Chinese mathematics. The library at the Tsinghua University in Beijing offers access to these holdings. In the USA the Cornell University Library has a good tradition in digitizing mathematical publications. At present project Euclid is active at both levels, the retrodigitization and the digital production of mathematical articles. In addition to this some journals spend a part of their income on digitizing their older volumes.

Taking all these developments into account, the content of the WDML is growing permanently like a patchwork and the links from Zentralblatt MATH to these holdings provide a first version of an integrated access to the content. Here the free offer of Zentralblatt MATH to look at the first three items of a hit is important for those who do not subscribe to Zentralblatt.

#### 4. EMANI and the ELibM

First steps towards a good cooperation between some of the stakeholders for the WDML could be observed in connection with the EMANI project. Partners of this group are SUB Göttingen, the Cornell University Library, the Tsinghua University Library, NUMDAM in Grenoble and Springer-Verlag. The acronym EMANI stands for Electronic Mathematics Archiving Network Initiative and was coordinated in the recent past by the author of this article.

Fig. 14: Entry page of EMANI

The main current activity is an NSF/DFG project for developing an integrated access to the distributed digital archives. During the project phase this refers to the archives at the three library partners, at NUMDAM, at Springer and to the electronic library in EMIS. In addition to this EMANI discusses possibilities for the long-term preservation and readability of digital content in mathematics. At present the aim of EMANI is to develop tools for the management of the long-term preservation in a distributed system. How to preserve the documents is left to procedures which have been established at other sites meanwhile.

An essential part of the WDML consists of digitally born publications, where in most cases the web offer still is accompanied by a printed version. In contrast to the early nineties, current submissions of articles

for publication do not rely on typesetting anymore. The manuscript has been prepared using a computer and the text is encoded in most cases in some TeX dialect. Up to some formatting everything for the posting of an article on the web has been prepared by the authors already. As a consequence only very few mathematical journals do not provide an electronic version.

To increase the visibility of their journals on the web and to gain advantages for the marketing of their products, publishers have bundled their journals and developed particular access structures for their bundles. This reorganizes the patchwork a little. Single journals published by a small publisher, a society or an academic institutions are sitting between the patches and loose visibility. It does not help them very much that digital production allows them to keep the prices low. This danger for the isolated journals had been taken into account in the early nineties already, when a European initiative lead to the installation of EMIS (European Mathematical Information Service, [www.emis.de/](http://www.emis.de/)) and on the American the Cornell University Library designed the Project Euclid. Euclid is hosting several of these low-budget isolated journals.

EMIS is based on a cooperation between FIZ Karlsruhe and the EMS. To deal with the problem mentioned above EMIS maintains and provides access to the Electronic Library of Mathematics ELibM. In contrast to Euclid it is required that the parts of the journals posted by EMIS are freely accessible. This may be combined with a moving wall solution, i.e., open access is provided only after the articles have become older than some years to be fixed for the journal case by case.

EMIS helps the journals to install an electronic offer in particular cases, but generally prefers to host a journal after it has been installed at the editors' website. It forwards the content to more than 40 mirror sites distributed all over the world. This kind of bundling and wide distribution gives the journals and other publications stored in the ELibM good visibility and facilitates access for potential readers world-wide. After a start with 10 journals only the ELibM became a rather popular choice for low-budget journals and free web offers of peer-reviewed journals. The current content of the ELibM counts on almost 100 active journals and some

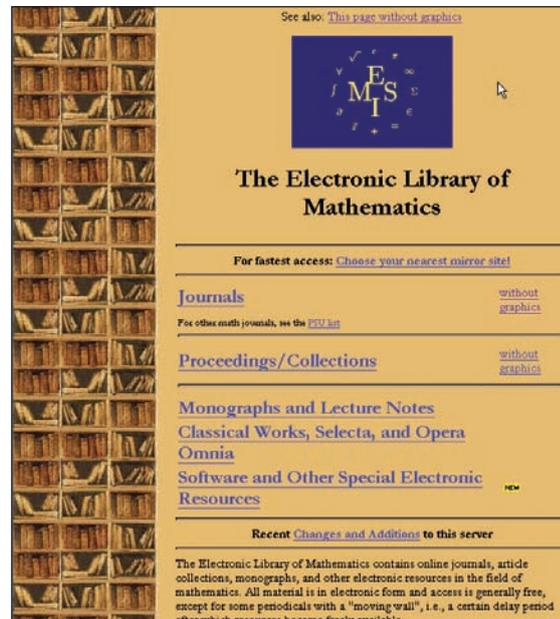


Fig.15: ELibM entry page

archived offers where the editors decided to finish the open access. About 15 gigabytes of storage space are needed for this. Integrated access is subject of an ongoing project and will be available soon. The close neighbourhood to Zentralblatt MATH guarantees quick linking and optimal handling by the reviewing service.

As refers to the content the WDML is in good progress. But the offers are quite heterogeneous if it comes to the document and metadata structure and to the business model for the different patches. Hence we are still far from having the ideal situation which includes integrated access. Spending some efforts on the user's side we can access a lot of articles, which were difficult to find in the era of printed publications. The linking with reference data bases like Zentralblatt MATH will help a lot, as long as there will be a free component for using this service. Possibly we have to be content with that solution on the global scale, which is not so bad at all.

Prof. Dr. Bernd Wegner  
Editor in Chief, MathEduc and Zentralblatt MATH