DML AND RUSDML – VIRTUAL LIBRARY INITIATIVES FOR COVERING ALL MATHEMATICS ELECTRONICALLY

Bernd Wegner

Abstract: With the rapidly growing activities in electronic publishing ideas came up to install global repositories which deal with three mainstreams in this enterprise: storing the electronic material currently available, pursuing projects to solve the archiving problem for this material with the ambition to preserve the content in readable form for future generations, and to capture the printed literature in digital versions providing good access and search facilities for the readers. Long-term availability of published research articles in mathematics and easy access to them is a strong need for researchers working with mathematics. Hence in this domain some pioneering projects have been established addressing the above mentioned problems.

Keywords: DLM, EMANI, ERAM, RusDML

Introduction

The paper will give a short state of the art report on some of these activities on the European level and the world wide plan to develop a global Digital Library in Mathematics (DLM). For example, in the archiving area as a special project for mathematics the Electronic Mathematics Archives Network Initiative (EMANI) had been designed. Having in mind that a distributed architecture would be more suitable and reduce the load on the partners for such a project, a network is proposed, which also might be a more open approach for extending the project from a initially restricted solution to a more comprehensive enterprise.

The Electronic Research Archive in Mathematics (ERAM) is a DFG-funded German project dealing with capturing the content of a classical bibliographic service in mathematics in a database, and combining this with the retro-digitisation of selected mathematical publications. This is extended now by further projects which shall try to retro-digitise the national mathematical heritage in several countries world-wide. In particular ideas to cover the Russian publications in a digital repository called RusDLM have been implemented in a project funded by DFG and RFBR. As further digitisation projects the French activity NUMDAM, pursued by Cellule MathDoc in Grenoble, and the European Cooperation in DIEPER have to be mentioned.

1. Electronic Offers and Their Providers

The impact of electronic devices on the daily life of researchers, teachers or other professionals results from a variety of tools and offers installed in local machines or made accessible through the internet. The part libraries are mostly involved in consists of electronic publications, or better electronic versions of printed publications. Some libraries already developed digital repositories containing retro-digitised publications, which had been obtained by scanning printed articles and books. But also offers which could be published only in electronic form become more and more important. I addition to this researchers and teachers increasingly take advantage of computer algebra systems and other computing software, and visualisation techniques using graphics software and image processing tools have become background for most of their presentations and publications. Finally, we should not forget that the internet has been used to establish a communication infrastructure which strongly facilitates their daily work and extend the possibilities for co-operation at distributed sites.

There is a wide range of providers of these offers, going from commercial publishers and learned societies to volunteers and single authors. Also the list of distributors and information brokers is a long one: libraries, databases and indexing services, internet-portals of different types, web browsers et al. In contrast to the "old world" of printed publications these providers have different aims and it is not always clear for the user what he really could expect from these services, when he is searching for some information or article of his own interest. Clearly, libraries try to transfer their system, they have developed for their printed holdings, to these new publications, and hence they still seem to be the most reliable information provider also with respect to electronic offers. But this role has to be acknowledged more widely and the offer has to be improved.

There are good reasons why libraries will be able to maintain their central role for distribution and storage of scientific information and succeed to extend this to the electronic media. They have developed precise and reliable access structures. Their service is free for their specific group of users, and this group is a large one in most cases. Even for external users they developed a good network of exchange facilities, which enables scientists to make their work really accessible for a wide community of users and to read the work of their colleagues without being confronted with bigger commercial barriers. Commonly libraries cover a broad area of subjects and within that they try to be relatively comprehensive. Independent from the frequency of their usage these holdings had been preserved and kept accessible with great care. The objectives of science libraries are user-oriented on one side. and on the other side libraries feel obliged to protect the treasure of knowledge they have accumulated in their collection. This makes them also the best choice for solving the problem of the long-term preservation of electronic publications.

Mathematics is a science where the availability of electronic publications and retro-digitised documents lead to a considerable improvements of the conditions for research. Hence, though some of the subsequent arguments may apply to all sciences, they turn out to be of particular importance for mathematics: Mathematicians and professionals applying mathematics need quick, reliable and integrated access to mathematical publications. Long-term availability of publications is a particular need in mathematics. Digitising of print-only publications and the adjustment of these offers to the current facilities provided for electronic publications leads to a additional series of problems to be solved. Electronic publishing offers a variety of additional information in mathematics which may be integrated into the access and display structures enhancing the traditional types of publications.

2. Some Evidence by Numbers

For non-mathematicians it is not clear at all that mathematics is so much different from other sciences as far as easy availability of older publications will be concerned. For some it is even hard to understand the subjects of mathematical research and the special way how this research is published. For example, extensions and improvements of older results only care about the publication of the additional achievements, and there detailed proofs are essential. Older results may and should be cited, but it is not honest to repeat their proofs in research publications, even if the understanding of these proofs is essential for seeing what the new results are about. Many proofs can be found at one place only. Hence an article is just an addition to a sequence of other articles, more or less tightly interrelated in a structure which combinatorically is more complex than a tree. It provides another shell to a core of theorems, propositions, examples, models and proofs representing the current knowledge of a subject domain in mathematics. Mathematical research articles commonly are rather thin, and the publication frequency of a mathematician is rather low compared to other sciences.

Admittedly, parts of such a domain may be exhibited comprehensively in monographs, but as can be seen by the variety of material in the research surveys in mathematics which have been published by VINITI (Moscow), for example, such monographs with detailed exhibitions of arguments only can cover a part of the domain of reference, giving a motivating introduction with proofs, while the surveys have no space to provide proofs at all, if they really want to be comprehensive. This underlines that references in mathematical papers are not just a matter of honesty, but that at least a part of them plays an important role for a complete understanding of the content of an article. Hence the following figures give a good evidence for the need to have also older mathematical publications available.

The evidence will be demonstrated in the case of three journals where the numbers are taken from an investigation by Joachim Heinze [*see Joachim Heinze*]. The most surprising figures (also to mathematicians) are the numbers of citations before 1992. In the case of the most traditional mathematical journal from North-America, the Annals of Mathematics, 60% of the citations in the 35 articles published in that journal in 2001 had a publication date before 1992. Vice versa, the number of cites from the volumes of 500 journals published in 2001 to the Annals was about 4.500 and 82% of them were before 1992. Looking at one of the first journals which published mathematics only (in contrast to journals which deal with several sciences), the Journal fuer die Reine und Angewandte Mathematik, founded as Crelles Journal in 1826, the first figure was 61% and the second 65%. Finally, these numbers still were high for a more "modern" journal which had been founded in the second half of the 20th century, the Inventiones Mathematicae: the first figure was 55% and the second one 68%. Such high

numbers of older citations are not common for most of the other sciences. It would be quite interesting to have a more comprehensive comparison of this type.

3. Current and Future Problems

In the "paper world" the long-term preservation of publications was simple on the first view, though at a closer look a lot of problems had to be handled. They mainly came from the deterioration of the paper or the binding of a book or journal, and they appeared after a comparatively long period in which the physical situation of the document could be considered as stable. Also a wide distribution of documents to several locations world-wide was a factor of stability, protecting them against being all destroyed simultaneously by the impact of wars etc.

For digital publications this period of stability turned out to be extremely small. What everybody experiences with his old releases of word-files, became true meanwhile for the readers of PDF-files, for example. Without conversions, if they exist at all, or simultaneous installation of several versions of the Acrobat-reader a whole range of PDF-files over the period, where the Acrobat reader was offered, was not readable anymore recently. This admittedly was a temporary problem, because the next release of the Acrobat-reader was capable to handle the full range of previous productions. But nobody can guarantee that a similar bug will occur again in the future.

This is only one problem. Another one is the stability of the physical carrier, where the data are stored, and there is a variety of plug-ins which depend on additional software to be offered with the electronic document. Current releases of this software may have a short life-time. What should we do with the document afterwards?

To solve this problem will be even more complicated when documents in mathematics are considered, because they are most likely to have software depending enhancements. Interactive documents will play an important role in the future. Furthermore, projects like MoWGLI [A. Asperti; B. Wegner] will develop different types of structures enabling semantic mark-up of documents. Hence preservation will go far beyond caring about the displayed text only. Structures, links and other informational background provided with electronic articles will have to be taken care of, and all these tools are in permanent evolution.

Hence the problem only can be attacked by a long-term approach as it is described with in EMANI in the next section.

4. The EMANI Project

There is a period of approximately 10 years during which electronic publications in mathematics developed from some offers in pioneering freely accessible journals to a first class publication facility with enhanced services in comparison to traditional printed publications. As mentioned above, older publications are still of big value for research in mathematics. Hence retrospective digitisation projects increased the current digital content in mathematics considerably. One major of these projects is ERAM (see *[H. Becker, B.Wegner]* or *[B.Wegner, ERAM*]) which will be described later on.

In the first half of 2001, the Electronic Mathematics Archives Network Initiative (EMANI) had been founded as a special project to develop models for the archiving of electronic contents in mathematics. Having in mind that a distributed architecture would be more suitable and reduce the load on the partners for such a project, a network is proposed, which also might be a more open approach for extending the project from a initial restricted solution to a more comprehensive enterprise. The initiative has been formalised in July 2002 at their workshop at Cornell University with the partners mentioned below as the first set of members and the author of this article as the co-ordinator of the project.

Thus, for the core of the network, a co-operational system of reference libraries and content providers like publishers and editors has been set up. In the ideal final version they are supposed to serve for a long list purposes: The basic action will be to store the digital content in mathematics from the content providers at the reference libraries. This will be complemented by retro-digitising all printed publications in mathematics from the content providers at the reference libraries, covering a big part of the publications in mathematics by electronic versions finally. On this basis first measures can be undertaken to care about the long-term preservation of this content in readable form. First projects for the technical support of this co-operation have been just initiated.

For example, to have the content stored somewhere will not be sufficient. Retrospective digitisation may lead to scanned images only, which hopefully can be accessed in some repository. As an important enhancement it will be necessary to improve the usability of the retro-digitised publications by introducing advanced linking and

searching facilities and to provide convenient and affordable access to the stored content for mathematicians and professionals using mathematics world-wide.

The reference libraries even may serve as a reference system for other libraries which want to store and provide part of the content or refresh their existing offers by updated material. Having in mind the long time scale of the publications provided through the network, going from articles from the 19th century to current publications, a system of distribution agents will be needed. This may be a good reason to develop new business models for a distribution of mathematical publications in a combined enterprise between reference libraries and content providers. But there is not only a theoretical discussion about potential activities in the future.

5. The Starting Point of EMANI

It will be reasonable to start with such a complicated enterprise only on a smaller well-controllable scale at first. Once the architecture and the action plan will have been made sufficiently precise, an extension may be considered. The current partners who collaborate for the first steps in order to implement the initiative on the side of the libraries are:

- The Cornell University Library, Ithaca, N.Y.: They have a good tradition in retrospective digitisation projects and are involved in the archiving discussion for other sciences also. In particular they are building up an offer of a bundle of electronic journals in mathematics through project Euclid. They serve as a mirror site for EMIS (see *[Bernd Wegner, ELibM]*).
- The State and University Library Goettingen: Also there some important retrospective digitisation projects like ERAM (see [*H. Becker, B.Wegner*] or [*B.Wegner, ERAM*]) and DIEPER are pursued. In addition to this the SUB Goettingen is obliged to collect all publications in mathematics. In this role they have a high reputation as a centre for access to mathematical publications. Moreover they also serve as a mirror site for EMIS.
- The Tsinghua University Library, Beijing: This library has experience with the digitisation of Chinese publications. They are a Chinese centre of excellence for installing and offering electronic publications.
- The Orsay Mathematical Library, Paris, in co-operation with the Cellule MathDoc in Grenoble: The group in Orsay is co-ordinating a quite comprehensive consortium of French mathematical libraries. The strength of the partner in Grenoble consists in their excellent retro-digitisation project NUMDAM ([*P. Berard*]).

The content providers for the start are the publishers running under the group lead by Springer-SBM currently and the electronic library ELibM offered through EMIS, the European Mathematical Information Service (http://www.emis.de). At the beginning of 2004 Springer-SBM started to digitise all periodicals published by this group, independent from the subject. But the pioneering offer of several of the best journals in mathematics in EMANI could be preserved. In contrast to this the ElibM is a co-operation of several journals and editors on a voluntary basis bundling electronic offers in a world-wide system of WWW-servers (see [*B. Wegner, ERAM*]). They agree to provide open access in general.

An important step in the first phase of the initiative consists of the stepwise transfer of the available electronic content from the content providers to the reference libraries. There it will be checked if the files still can be used for the archiving, adjustments will be made in the case of files which are unsuitable for this and recommendations will be developed how the content providers could care about a more convenient delivery in future cases. Also new archiving related meta-data have to be defined, and an integrated access structure satisfying the needs of all kind of experts who want to work with the archive will be one of the central achievements of the further work in the future. Though links from reference databases could satisfy many of the needs of the mathematicians to get access, the professional handling of the archives will require more than just their meta-data.

To check the TEX-files for their usability without any appropriate system in the background turned out to be a tedious task. This was never considered as a part of the work of librarians, if we think of the period where TEX only was considered as a tool for preparing beautiful camera-ready manuscripts. But taking TEX as a tool for the mark-up of publications on the ASCII-level and providing files, which are most suitable for long-term preservation, the EMANI-partner Goettingen reacted very quickly, by establishing a project to develop tools for an automatic checking of the usability of TEX-files. A prototype for the tool could be presented in 2004.

6. ERAM – Combining a Database with an Archive

Also for older documents searchability will be an important requirement to enable the researcher to find his way in the huge knowledge base of mathematical achievements. Admittedly, no current search engine is able to locate a statement in its abstract meaning. Names for some of them will help, and classification codes of special subject areas will restrict the set of documents where to look for the desired information considerably. Hence literature databases for the classical period of mathematics are desirable. They should offer the same facilities like the current literature information services in mathematics, and even more, they should also provide links to the future given by modern mathematics. This is the starting point for the project ERAM which also will be called the Jahrbuch-project for short.

The acronym ERAM stands for "Electronic Research Archive for Mathematics". The project is funded by the Deutsche Forschungsgemeinschaft (DFG). The aim of the project is the installation of a (digital) archive of articles relevant for mathematical research, full searchability and access through a database, captured from the "Jahrbuch ueber die Fortschritte der Mathematik" (1868-1943). The most comprehensive current literature database in mathematics, Zentralblatt MATH, was founded at the end of the Jahrbuch period. The first step of the ERAM-project is the production of a bibliographic database, the JFM-database, capturing the content of the Jahrbuch ueber die Fortschritte der Mathematik (JFM). This has been finalized in a first version in the first half of 2003. Modern literature databases provide several search options for which the information could not easily be extracted from the text of the JFM. Hence, editorial enhancements are under preparation, and moreover historical links are provided to modern research as far as possible. The only formalised subject information in the JFM consists of the subject headings which are stored in the database like a raw classification. A more precise description of their subjects will be obtained by additional intellectual indexing work. The corresponding experts provide an English translation of the title of the single document, they add a subject classification according to the MSC2000 scheme and assign some English keywords.

All data from the Jahrbuch have been keyboarded now. They are made accessible in this form in the web, and though for many items the enhancements are still missing the database has found a lot of grateful users. As a consequence combined searches with the database Zentralblatt MATH are offered. In addition to its usage as high-quality source for information on classical mathematics, the JFM-database will provide access to a digital archive to be built up within the project. For this selected publications are scanned (as gif-images) and stored in a document management system. Currently there are no conversions of the images into text files. To allow text searches in the archive, text files will be an important addition to the scanned images. But the generation of these data will be a matter of a later phase of the project. Conversion programmes have been improved considerably, as had been demonstrated at a satellite meeting to the ECM 4 in Stockholm, and they able to tackle the problems which occur with formulas in mathematical texts. A first step in this direction is made by a project based on the co-operation of experts from Japan, Germany and the United States (see [*G. Michler*]). Bt this has been topped by a conversion method provided by Tim Dokchitser meanwhile.

The scanned material includes journals like Mathematische Annalen, Mathematische Zeitschrift, Inventiones Mathematicae, Commentarii Mathematici Helvetici, for example. The Journal fuer die Reine und Angewandte Mathematik will be added at the end of 2004. Most of the journals which have installed recent electronic versions in EMIS (European Mathematical Information Service) agreed that all of their print-only back volumes could be digitised and offered within ERAM, and this also has been done. In ERAM, about 1 million pages have been scanned so far, and the capacity of the project will be sufficient to go for about 1.2 million pages. For more details see the references [*H. Becker, B.Wegner*] and [*B.Wegner, ERAM*], or the ERAM-homepage under http://www.emis.de/projects/ clicking on the box for the Jahrbuch.

7. The Global Digital Mathematics Library – DML

ERAM could be considered as a part of a global initiative to have all mathematics digitally available. It has a lot of overlap with EMANI and both projects a tightly linked with each other. But in contrast to EMANI the global initiative at first will concentrate on retro-digitization, i.e. the preparation of digital versions of texts which are not yet digitally available. Long-term preservation is a secondary aspect of the DML at present. Clearly, in addition to ERAM there are several other digitization projects on the way, general projects like JSTOR, DIEPER, and the Elsevier backfiles system, and projects in mathematics like NUMDAM [*P. Berard*] or the national heritage activity

in Colombia by Victo Albis [*V. Albis*]. The Tsinghua University Library succeeded to digitise more than 50 Chinese journals. The digital offers made accessible through the EMANI homepage comprise more than 100 journals. At KISTI in Korea 16 mathematical journals have been digitised. All this has to be taken into account for getting an impression about the state of the DML.

In 2001 John Ewing prepared his White Paper [*J. Ewing*] in which a rough estimate has been made how much money would be needed to develop global digital mathematics library (DML) containing all mathematics in digital form. This estimate was in the order of 100 million US Dollars. But that was not the main achievement of that paper. It contained a lot of structural considerations for such a library, and it also addressed the immense problems we will be confronted with when we really want to pursue such a project. As a caveat when reading this paper, one should be aware that it describes an ideal solution, and some parts like a central repository (by intention) do not reflect very well what has been developed already. For example, at present only a system of distributed repositories could be imagined, because proprieties and aspects of cultural heritage have to be respected. Furthermore, a distributed system can hook on existing providers like libraries, and this will be more efficient than the installation of an extra infrastructure to manage the DLM, as far as the costs will be concerned.

As a consequence a planning grant had been given to Cornell University by the National Science Foundation of the U.S.A. to make a feasibility study for the DML. This will be done during two workshops, where the first one took place already in Washington D.C. at the end of July 2002, the second one at the SUB Goettingen in May 2003. The 25 participants from different kinds of institutions set up a scheme to develop a plan for the DML. An initiative has been formalised, working groups have been designed and a Steering Committee has been chosen to guide the progress of the discussion during the next future. More or less the scheme reflects a part of the project administration for EMANI, and indeed DML may profit a lot from the preparations in EMANI.

It will the subject of an article of its own to go into all the details having been addressed by the working groups, but one of them should be explained here, because it is basic for the definition of the global project as well as for the description of the environment for similar national or local projects.

How can we determine what has to be considered as the content of mathematics?

Talking to mathematicians it will be noticed rather soon that the idea what should be covered by the DLM is quite vague. There are ongoing projects which have selected items for retrodigitization according to different aspects. These patches of the global DLM can be defined easily, but they cannot serve as a model for a comprehensive coverage of mathematical publications. Hence some more concrete questions arise naturally:

- Do we really have the chance and the interest to cover all mathematical publications world-wide by the DLM? If not so, the selection criteria have to be discussed. But also in the other case we have to decide on selection criteria, because not everything could be done immediately and a time schedule for building up the DLM step by step requires an order and hence a selection.
- Which publications may be considered as a part of mathematics according to subject area?

It will need a lot of efforts and patience to arrange such a list of contents and somebody has to administrate this. People are most likely to escape from this by deciding not to care about such a list at all and digitise what will be just in their mind or easily available. This is good for the patchwork, but it will ruin the global idea. To work on the global solution, four dimensions have to be considered: When do we start and how far back should we go? How much mathematics is supposed to be in the document? There are impact on research, potential user interest in having the document available, depending also on different user communities (research, education, applications, history etc.), quality, availability. Where should be the priorities? There is a geographical dimension which may be associated with priorities for the DLM-actions. How should the DLM project spread out from current initial activities covering content from all over the world ?

But there is also a cultural dimension. Though the global approach is a challenging idea, the development of the repositories should take national interests and funding possibilities into account. Hence distributing the content to single projects has to respect what had been covered already and what should remain under the guidance of a special mathematical community. Only the remaining content may be open for adoption for retro-digitisation. To distinguish this will be one of the main tasks during the content determination and it will be a delicate task, because very quickly there may be the impression that one party wants to buy out the mathematical heritage from another one.

The activities of the Planning Group have been finalized after the second meeting in Goettingen. The result will be a report to be presented to the NSF and to be made public by the chief researchers of this group. To go on with the coordination activities a committee has been installed by the IMU. But this is only one option to keep the DML initiative going. The DML-EU application is another integrating activity, which keeps the different parties talking to each other and promotes to think about the installation of digitisation projects being funded by other parties. One reason for the is that funding from the FP6 only can be used for networking and research activities. Basic digitisation only can be funded in the form of small test beds. Mass digitisation will have to ask for funding from other sources.

8. Aims and Mission of RusDML

The initial goal of RusDML (Russian Digital Mathematics Library) is to digitise a core collection of Russian journals in mathematics, which so far have been available in printed form only and, by making them accessible in the web, to facilitate the world wide access to them. Having succeeded with this a further activity may go for comprehensiveness, i.e., to perform the digitisation of all Russian mathematical publications, including monographs, series of collections of papers, encyclopaedias, handbooks, proceedings volumes, deposited articles etc. The project has started in the middle of 2004 and is funded by DFG and RFBR. The general need to establish such a project is the same a for the DML.

RusDML is planned in accordance with the basic requirements for the DML: The archive should be open and accessible world-wide. Distributed copies should guarantee the safety of the data and facilitate the access from different parts of the user community. DLM should be established as global network providing access to interlinked digital publications in mathematics. As a result of the DIEPER project a first sample issue for RusDML even will be available in advance to the project itself. The contents of the most traditional Russian journal in mathematics, Mat. Sbornik, had been scanned by the DIEPER partner in Helsinki, and after some additional work on the access data this journal will be available as a RusDML prototype due to kind agreements with the Helsinki University Library for using their files and the Russian Academy of Science (RAS), Moscow Branch, to make the digitised articles freely accessible through the web.

As a key issue the Russian-German cooperation between several partners in both countries will be the organizational base of the project. Scientifically this is a consequence of the traditional good cooperation between Russian and German mathematicians for three centuries. As everybody knows, this cooperation survived some political catastrophes. But even now, when we have a period where Russian mathematicians partially try to publish in other languages, there is a comparatively high interest in Russian publications in Germany and other European countries. It is no question that for Russian mathematicians the digital offer to be installed with RusDML will be a highly desirable improvement of their literature supply. But the two libraries involved on the German side still have the image to be reliable and comprehensive reference sites for this. Providing the content of RusDML will make them unique sites for users who are not likely to go to a provider in Russia. As a consequence, a bilingual access structure with enhanced facilities for those with weak Russian reading capabilities will be one of the most important requirements for RusDML. The conceptual and technical background for this also will be a pioneering tool for digital offers of other Russian publications.

The main Russian partner is the Russian National Public library for Science and Technology (GPNTB). On the level of scientific advise for the project the advice of the Mathematics Division of RAS will be mportant. Interests of other Russian libraries like those from Kazan, or MGU, or RAS will be respected in bilateral agreements. RusDML will establish GPNTB as the centre of excellence for digital offers of Russian mathematics. But in spite of this there is no aim to interfere with the interests of other mathematical libraries in Russia. Hence their aims and mission will be respected and taken into accordance, when delivery of documents, linking of offers and mirroring services should be taken into account. GPNTB is the major library in Science and Technology for the Russian Federation. Its collection comprises 8 million items of national and foreign publications. The library provides comprehensive access to Russian collections in its role as a State Depositary and recipient of obligatory free of charge copies (mandatory copies) of all publications in their domain. All journals selected for RusDML are available at the collection of GPNTB, starting just from the first issue of the journal in its first publication year to the current production. Moreover, the library is experienced in library automation and information technologies. The Russian Academy of Sciences organizes born electronic offers of their journals and provides them freely for Russian users through their IZIR system. Like with Mat. Sbornik they will consider the digitisation of their journals

as an added value, and customize everything in a convenient way for their users. In this sense RAS clearly supports the RusDML initiative and this will be an extremely helpful assistance for the negotiations with other Russian publishers and editors for getting the license to digitise their materials. Concerning the scientific exploitation of the RusDML RAS may play a leading role to explain the many advantages of the enhanced digital offers.

9. The German Part in RusDML

The German partners are the State and University library in Goettingen (SUB), the Technical Information Library in Hannover (TIB), and the Technical University Berlin (TUB) representing the contributions from Zentralblatt MATH through its editor-in-chief given by the second author. The different roles of these partners will be explained below. But as an essential common facility it had been agreed, that all three libraries, GPNTB, SUB and TIB, have the option and almost the obligation to install a full copy of RusDML. All project participants are well prepared for the collaboration, because they have pretty good collections in mathematics and they have long experience with the handling and administration of electronic offers.

Zentralblatt MATH has a very special role in this cooperation . Its main duty is to provide a comprehensive reference data base in mathematics. It provides bibliographical data, indexing information and reviews or abstracts in English. Hence the core metadata for RusDML will be available there, because all journals in RusDML are evaluated by Zentralblatt, and employing the linking facilities from the database to full text offers, it can be used as a simple access tool to the holdings of RusDML. The idea is to integrate the reviews as a special addition into the metadata. Hence also users with low reading capability in Russian can decide, if they really want to go into the details of an article or not.

As mentioned above, there are some basic requirements for the project. Most importantly the digital archive should be easily accessible world-wide. All three participating libraries should spend combined efforts to take care of the long-term preservation and readability of the digital collections. Later upgrades of the offers can be imagined leading to more convenient access to electronic archive and to improved search facilities. For example, one important item is the linking from the references to their web offer. To achieve these goals all partners supposed to share their efforts and results and as a consequence they should serve as mutual mirror sites for the complete archive of RusDML. The participating libraries will support international standards for RusDML as recommended by the DML project and others. This keeps the project open for cooperation with other initiatives to be capable for further expansion to cover the Russian publications more completely. There may be additional digital collections provided by other institutions, which are not in the core list of documents recommended for RusDML, for example.

With respect to content several stages of the project have to be considered. On the first stage RusDML will start with processing journals from a core list of about 120 titles, which are covered by bibliographical databases of Zentralblatt MATH and the Jahrbuch database. But in addition to this a big variety of Russian publications in mathematics is available. Hence the core list should be corrected and extended. For instance, in the collection of GPNTB some quite interesting mathematical journals can be found which are not listed as journals in the Zentralblatt database. For this a list of additions has been developed, possibly containing as a journal, which has been classified as a series of collections of articles by Zentralblatt. A registry of Russian publications in mathematics will be developed and extended during the project period.

Coming to the total amount of work in the first stage the following figures have to be considered. Starting with the approximately 120 Russian journals of the first list, which have been processed by Zentralblatt and the Jahrbuch, joint estimates by GPNTB and SUB came to a figure of about 2 million pages. Using the existing digitisation infrastructure at both sides it is agreed that the handling of the structural metadata, which are necessary for controlling the page numbers and the scanning are shared at equal parts between GPNTB and SUB. This delegates about 1 million pages to GPNTB. This will be done on the basis of uniform formats and common protocols with respect to technical issues.

Another nontrivial problem for the content part is the work on licences and permissions. To convince publishers and editors to make their printed publications available on the web needs a lot of promotional work. Some will agree immediately, others will have a lot of reservations and prefer to wait what happens with the first set of publications in RusDML. Thinking about contacting every author will make the work tedious and increase the efforts considerably. The copyright discussion is an open matter in this area, but librarians will have intermediate solutions to survive with an offer in the net in a more or less legal way.

10. Added Value of RusDML

Having established a digital version of a journal for its complete publication period will be a first step only. According to the work described above there will be a service related access and navigation structure enabling readers to browse the offer and to read (and print) the text of the articles.

The automatic generation of reference links is likely to operate very soon. The references could be distinguished from the other text in the image. Applying OCR to that part will be able to get structured text information from each reference. Flexible look up systems will contact reference databases like Zentralblatt MATH and arrange a search for exporting the identifier of the reference in order to add this to the metadata of the document under consideration. This can be used to arrange the links from that document to the complete text of the reference, if a digital version is available. Clearly, an adjustment of these tools will be required which can handle Cyrillic characters.

Going beyond these formal procedures dealing with the scanned image itself, the editors and mathematicians interested in the corresponding journal will have the possibility to enrich the information associated with that journal. The static sequence of images and metadata produced for a journal will be capable to store comments, historical remarks or any kind of addition, which seems to be of interest in relation to the scientific merits of the corresponding article. Hence there will be a dynamic aspect in the management of a journal turning even so retro-digitised part into a living archive. This input will be subject to the initiative and the control of the editors. It will provide an improved view of the location of the journal in information space and of its role in the development of mathematics.

This will bean add ed value for the journal, and it only can be obtained in a convenient way, after having the journal digitised, and equally important, after having provided a structure where useful additional information could be handled in a searchable way.

Bibliography

[Andrea Asperti; Bernd Wegner] MOWGLI – A new approach for the content description in digital documents. Ninth International Conference "Crimea 2002" Libraries and Associations in the Transient World: New Technologies and New Forms of Cooperation. Conference Proceedings. Sudak, Autonomous Republic of Crimea, Ukraine, June 8-16, 2002, Volume 1: 215-219.

[Bernd Wegner, ELibM] ELibM in EMIS – A Model for Distributed Low-Cost Electronic Publishing. Eight International Conference Crimea 2001O Libraries and Associations in the Transient World: New Technologies and New Forms of Cooperation. Conference Proceedings. Sudak, Autonomous Republic of Crimea, Ukraine, June 9-17, 2001, Volume 1: 317-320.

[Bernd Wegner, ERAM] ERAM – Digitalisation of Classical Mathematical Publications. Seventh International Conference Crimea 2000O Libraries and Associations in the Transient World: New Technologies and New Forms of Cooperation. Conference Proceedings. Sudak, Autonomous Republic of Crimea, Ukraine, June 3-11, 2000, Volume 1: 268-272.

[Galina A. Evstigneeva, Andrei I. Zemskov] RusDML – A Russian-German Project for Establishing a Digital Archive of the Russian Mathematical Publications, Lecture Notes in Computer Science, vol. 2730, Springer-Verlag, 2003

[Gerhard Michler] How to build a prototype for a distributed digital mathematics archive library. Proceedings MKM 2001, Linz, http://www.emis.de/proceedings/MKM2001/.

[Hans Becker, Bernd Wegner] ERAM – Digitization of Classical Mathematical

[Joachim Heinze] Presentation at the first EMANI workshop in Heidelberg, February 2002 (article to appear in the Proceedings of the EIC-Satellite Conference to the ICM 2002, Tsinghua University, Beijing)

[John Ewing]Twenty Centuries of Mathematics: Digitizing and disseminating the past mathematical literature. http://www.ams.org/ewing/Twenty_centuries.pdf

[Pierre Berard] Presentation at the San Diego DML-meeting, Joint Mathematics Meeting, January 2002 (see also http://www-mathdoc.ujf-grenoble.fr/NUMDAM/).

[Victor Albis] Conservacion del patrimonio matematico colombiano.
http://www.accefyn.org.co/historia-matematica/histmatcol.htm;
http://www.accefyn.org.co/proyecto/conservacion.htm; http://168.176.37.80/matepro.html
Publications, Proc. ECDL 2000, Lecture Notes in Computer Science 1923, 424-427 (2000).

Author Information

Bernd Wegner, Prof. Dr. – Fakultaet II, Institut fuer Mathematik, TU Berlin, Sekr. MA 8-1, Strasse des 17. Juni 135; D – 10623 Berlin, Germany; E-mail: wegner@math.tu-berlin.de