

Recent Progress with RusDML

Galina A. Evstigneeva Bernd Wegner

GPNTB Moscow, 12 Kuznetski Most, Moscow, Russia

e-mail: gae@gpntb.ru

Mathematisches Institut, Fakultät II, TU Berlin, Sekr. MA 8-1

Straße des 17. Juni 135, D – 10623 Berlin, Germany

e-mail: wegner@math.tu-berlin.de

Abstract. This is an update of the description of RusDML given in [3] and [4], where RusDML stands for Russian Digital Mathematical Library. The further work on the main features of the project like the professional structure for the digitisation process, collaborative structures for caring about the preparation of the bilingual metadata, international cooperation for caring about the input and providing a system of mirrors of the archive with several access possibilities are described as far as they have modified or extended in the period when the financing of the project had been secured. Several tools developed for the project will play a pioneering role and can be used for digitisation activities in other fields. They will be highlighted shortly again.

1. Objectives and basic requirements

The goal of RusDML is to digitise a core collection of Russian journals in mathematics, which are so far available in printed form only and, by making them accessible in the Web, to facilitate the world-wide access to them. This is in accordance with the global ambitions of the WDML (see the article [9] in these proceedings for recent developments and the White Paper [5] by J. Ewing for the initial phase).

The starting point for the WDML were the following arguments: Mathematics is a science where the availability of electronic publications and retro-digitised documents leads to 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, – mathematical results do not have a date of expiration. This gives evidence for the need and the actual demand to install RusDML.

In addition to better access, digital mathematical archives would serve as electronic repositories and would provide a preservation facility for the printed collections. In this line one could see also a rational exploitation of library premises, providing better conditions for preservation especially for acid-paper publications, which are endangered by deterioration.

The basic requirements for RusDML in its core solution are the following: The archive will be open and accessible world-wide. Distributed copies will guarantee the safety of the data and facilitate the access from different parts of the user community. RusDML will be part of the global network providing access to digital publications in mathematics. It will support the needs of Russian libraries other than the project partners to install local access.

Referring to other digitisation activities on the academic level like ERAM (see [1], [8]), standing for Electronic Research Archive in Mathematics, and NUMDAM (see [7]) the development of RusDML will integrate their achievements as far as possible. Links with other offers will be highly desirable in a later phase. For example, as soon as the translations of the main Russian mathematical journals will have been digitised, the installation of mutual links between the translation and the Russian original is scheduled. This has been agreed with the publishers of the translation journals. For a more comprehensive account of what has been digitised so far the reader is referred to [9]. Furthermore, the retrospectively digitised content will be matched with the offer of digitally born material, enlarging existing collections of full text electronic documents.

2. The co-operational network

As a key issue the Russian-German cooperation between several partners in both countries will be the organizational base of the project. Though we are experiencing a period where Russian mathematicians partially try to publish in other languages, there is a comparatively high demand of Russian publications outside Russia. The two libraries involved on the German side, SUB Göttingen and TIB Hannover have the image to be reliable reference sites for mathematical publications. Providing the content of RusDML makes them unique sites for users outside Russia who are unlikely to go to a provider in Russia. For Russian mathematicians the digital offer to be installed with RusDML will be a highly desirable improvement of their literature supply. As a consequence, a bilingual access structure with enhanced facilities for those with weak Russian reading capabilities is pursued as one of the most important services of RusDML. This feature also represents a pioneering work for other digital offers of Russian publications.

The main Russian partner for RusDML is the Russian National Public library for Science and Technology (GPNTB) in Moscow (see [6]) scientifically backed up by the Mathematics Division of RAS. Interests of other Russian libraries will be respected in bilateral agreements, because RusDML may establish GPNTB as one centre of excellence for digital offers of Russian mathematics. After the Deutsche Forschungsgemeinschaft (DFG) has decided in the first half of 2004 to fund the German part of the project, GPNTB has been provided with funds from RFBR to care about the Russian part.

Being the Russian backbone of RusDML GPNTB has no ambitions 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. Some first agreement with the Library at the Steklov Institute of Mathematics in Moscow has been arranged already. They are coordinating a network of mathematics libraries in Russia with the aim to improve the access to digital mathematical literature in particular. Clearly they have high interest in a copy of the content of RusDML. On the other side they already started with digitising two Russian journals themselves. Hence as a natural consequence of the general aims and mission of RusDML it was agreed that they give copies of the journals digitised at their site to RusDML and receive a copy of the complete system for their purposes.

The German partners are the State and University Library in Göttingen (SUB), the Technical Information Library in Hannover (TIB), and the Technical University Berlin (TUB). The different roles of these partners have been explained in detail in [4]. The funding on the German side is attributed to the partners according to their role in the project. For example, the TU Berlin will be in charge of the delivery of the metadata in English for the articles to be digitised. They will be taken from Zentralblatt MATH in accordance with the editor-in-chief of that service. But the main Russian journals have a comparatively long lifetime already. Hence the metadata need some standardisation and some improvements before being integrated in RusDML. For this purpose a part time position of a scientific editor is assigned to TU Berlin.

Zentralblatt MATH provides a comprehensive reference data base in mathematics, including bibliographical data, indexing information and reviews or abstracts in English. The core metadata for RusDML are available from Zentralblatt, because all journals in RusDML are indexed by Zentralblatt. 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. Providing the reviews as a special addition to the metadata, also users with low reading ability in Russian can decide, if they really want to go into the details of an article or not. To have the quick access available as soon as possible represents a further argument for integrating this facility into RusDML.

3. Content of RusDML

A big variety of Russian publications in mathematics is available. Hence several stages for the development of a comprehensive archive 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. This core list is open for modifications whenever an important publication to be covered by RusDML is detected. It is taken as a first release of a registry of Russian publications in mathematics. It includes more journals than the current Russian ones. There are journals, which were published in the USSR and after the change for the time being are published by a publisher in the New Independent States. Furthermore there are titles, for which the publication has been cancelled, or it has been temporarily interrupted and then taken up later again, possibly under a new name. This has to be taken into account for taking care of the registry and determining the content of RusDML.

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.

Furthermore, the problem of the integration of externally produced content has to be taken into account. The part to be imported from the library of the Steklov Institute of Mathematics already had been mentioned. There are additional digitisation activities at the library of the Kazan University. They will cover more or less the mathematical heritage of that region as far as journals will be concerned. To make that content compatible with what has been produced within RusDML requires both, the production of metadata in full accordance with the scheme developed for RusDML and the determination of the interfaces for importing this material into the workflow of RusDML or into the archive itself.

The current work in RusDML deals with three journals. The contents of the most traditional Russian journal in mathematics, *Matematicheskij Sbornik*, had been scanned by the DIEPER partner in Helsinki. The Helsinki University Library gave the approval that SUB Göttingen may use the files for the installation of a freely accessible offer in the Web. This was supported by a corresponding licence from the Russian Academy of Science (RAS) in Moscow, the owner of that journal. For a professional offer some additional work on the access data the scans of this journal could be made accessible. Before RusDML was funded, this work could have been done in ERAM, and some steps already were undertaken there. Now the production of the full metadata according to the scheme agreed for RusDML will be integrated into the workflow for that project, making available this journal for open access in the Web soon.

The other two journals are *Sibirskij Matematicheskij Zhurnal* and *Algebra i Logika*. The owner of these is the Sobolev Institute of Mathematics at the Russian Academy of Science in Novosibirsk. Their editors gave the approval to have the journals digitised and offered for open access in the Web without further restrictions and in very cooperative way. The digitally born version of *Sibirskij Matematicheskij Zhurnal* is freely available through EMIS already. When its digitisation will have been finalized a complete integrated digital version of that journal will be available in the Web.

4. Tools for RusDML

To start with the work for RusDML several agreements and tools had to be set up. Some of the standards had been taken from the recommendations for the WDML (see [2]) and EMANI, because the main digitisation centre for RusDML is located at SUB Göttingen and associated with both initiatives. But not all requirements for RusDML could be covered by these developments.

Since at least four partners are supposed to work at the documents for RusDML a workflow was developed in full detail in the second half of 2004. This includes facilities to deliver input by batch jobs, such that the system is not locked by online input from one of the partners. For the cooperation with Zentralblatt import and export formats had been set up. Some standardisation facilities have been arranged to identify journals' names and authors' name as far as possible.

For the production of metadata an application profile had been developed. This includes the decision, which metadata should be made available in Cyrillic. It also fixes the transliteration table. For indexation several systems may be used simultaneously. The MSC codes are available from Zentralblatt. But Russian users may wish to use UDKs, and librarians would like to see DDCs. Conversion tables are available for the transition between MSC and DDC. The same for UDK is still missing, but would be a great help for developing the UDKs within the project.

An open problem will be a combined search engine. The German partners have one for the metadata described in Roman letters, GPNTB has its own for the Cyrillic. There are several arguments for such a combined tool, but at present the development of the content has higher priority.

5. Upgrades

There are a lot of facilities for RusDML, which cannot be set up by the current project and have to be left to further projects. Most importantly the content should be increased later, including monographs and deposited articles for example. But that should be decided only after the core of RusDML has been established successfully.

Later upgrades of the offers should lead to more convenient access to the electronic archive and to improved search facilities. For example, one important item is the linking from the references to the Web offer of the cited papers (reference linking). Another added value for the service consists of the possibility to do full text searches. These options are implemented by other digitisation projects already, but in the given case the tools for scanned documents in Russian still have to be looked for.

Finally there also may be interest in a living project on the scientific level. The editors and scientists at RAS will have the possibility to enrich the information on their journal by comments, historical remarks or any kind of addition, which seems to be of interest in relation to the scientific merits of the corresponding article. This will be an added 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.

References

- [1] Becker, Hans; Wegner, Bernd: ERAM - *Digitisation of Classical Mathematical Publications*. Proc. ECDL 2000. Lecture Notes Comput. Sci. **1923** (2000), 424–427.
- [2] URL: <http://www.library/cornell.edu/dmlib/>
- [3] Evstigneeva, Galina A.; Wegner, Bernd: *O proekte sozdaniya elektronnogo archiva russkich publikatsii po matematiki*. Proceedings of LIBCOM 2002, Yershovo, November 2002.
- [4] Evstigneeva, Galina A.; Zemskov, Andrej: *RusDML – a Russian-German project for establishing a digital archive of the Russian mathematical publications*. Bai, Fengshan (ed.) et al., Electronic information and communication in mathematics. ICM 2002 international satellite conference, Beijing, China, August 29–31, 2002. Revised papers. Lect. Notes Comput. Sci. **2730**, 44–51, Springer, Berlin 2003. Zbl 1038.00011

- [5] Ewing, John: *Twenty Centuries of Mathematics: Digitising and disseminating the past mathematical literature*. http://www.ams.org/ewing/Twenty_centuries.pdf
- [6] URL: <http://www.gpntb.ru>
- [7] URL: <http://www-mathdoc.ujf-grenoble.fr/NUMDAM>
- [8] Wegner, Bernd: ERAM – *Digitalisation of Classical Mathematical Publications*. Seventh International Conference Crimea 2000. 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.
- [9] Wegner, Bernd: EMANI – *Leader and follower for the WDML*. See these Proceedings.

Received December 15, 2004