RusDML 2008
Current Facilities of the Core Archive of Digitized Russian Publications in Mathematics

Bernd Wegner

Mathematisches Institut, Fakultät II, TU Berlin
E-mail: wegner@math.tu-berlin.de

Abstract. The purpose of this extended abstract is to summarize the report on the final achievements of the RusDML project, where this acronym stands for Russian Digital Mathematical Library. The initial phases of the project have been described in [1] and [2]. Several main features have been installed meanwhile: a professional structure for the digitisation process, collaborative structures for caring about the preparation of the bilingual metadata, an international cooperation for caring about the input. A system of mirrors of the archive with convenient access facilities will be available soon after the digitization of all journals, for which digitization licences are available, will have been finished in July 2007. Some tools developed for the project are playing a pioneering role. Third parties use them already.

Key words: retrospective digitization, Russian mathematics, bilingual access

1 Objectives and Basic Requirements

The goal of the RusDML project is to digitise a core collection of Russian journals in mathematics, which so far were available in printed form only and, by making them accessible in the web, to facilitate the worldwide access to them. This is in accordance with the global ambitions of the WDML. The basic requirements for the digital content developed within RusDML are the following: The archive will be open and accessible worldwide. 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.

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. Furthermore, the retrospectively digitised content will be matched with the offer of digitally born material, enlarging existing collections of full text electronic documents.

2 Russian Mathematics and the Content of RusDML

As a key issue for a Russian-German project cooperation between several partners in both countries has to be taken as its organizational base. 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. Hence there is a good motivation for libraries to support this project. 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 are pursued as one of the most important services of RusDML. This feature also represents a pioneering work for other digital offers of Russian publications.

A big variety of Russian publications in mathematics are available. Hence several stages for the development of a comprehensive archive have to be considered. The RusDML project started with processing journals from a core list of about 120 titles, which are covered by the bibliographical databases Zentralblatt MATH and the Jahrbuch database. This core list is open for later modifications whenever an important publication to be covered by RusDML is detected. It may be 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.

Two problems lead to a restriction of the first release of RusDML. One was the problem to get digitization licences from Russian editors and publishers. The initially promising attitude of the Moscow Branch of the Russian Academy of Sciences had changed slightly during the initial phase of the project, making the inclusion of most of their journals impossible. Fortunately the branches in Novosibirsk and in St. Petersburg did not share this attitude. The other problem was the handling of the metadata, which slowed down the workflow considerably. This became more complicated than estimated initially.

3 The Co-Operational Network

The main Russian partner for RusDML is the Russian National Public Library for Science and Technology (GPNTB) in Moscow (see `http://www.gpntb.ru`). In order to avoid conflicts with other Russian libraries their interests will be respected in bilateral agreements. Some first agreement with the Library at the Steklov Institute of Mathematics in Moscow has been arranged already.

RusDML may establish GPNTB as one centre of excellence for digital offers of Russian mathematics. After the Deutsche Forschungsgemeinschaft (DFG) had 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.

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 funding on the German side is attributed to the partners according to their role in the project. The different roles of these partners have been explained in detail in [2]. Some of the activities will be described in the report again in order to give some impression of the efforts, which are needed for the production of the metadata.

Employing the linking facilities from the Zentralblatt database to full text offers can be used as a simple access tool to the holdings of RusDML. It was decided to include the reviews as a special addition in the metadata. This will enable users with low reading capability in Russian to 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.

To obtain the full set of metadata all partners will have to deliver some contribution. In addition to the above SUB has to take care of the structural metadata, needed for the digitization. The part of GPNTB is to enter the metadata in Cyrillic. Finally TIB will add the metadata in transliterated form in accordance with two schemes, ISO 9 and DIN 1460.

4   Tools and Standards 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. To benefit from the Zentralblatt data 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 tables. For indexation several systems like MSC or UDK may be used simultaneously.

A final tool 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. Initially the development of the content had higher priority. But now some first search facilities are available already, and there is a good chance that even text search in Cyrillic could be pursued until end of the project.

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.

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, which is on the way to be implemented soon. These options are available for other digitisation projects already.

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.

But all this has to step back behind efforts to increase the content on the journals level. At present we have a situation where the rights for digitization are reserved, open access is denied in many cases and the editors and publishers want to develop digitized offers on their own. But in contrast to Western commercial publishers no or only a little money could be spent for this at the Russian side. Hence it is very likely that offers with low-resolution scanning and scarce metadata will be installed. The results of RusDML will be a measure of quality for this and by this hopefully prevent that the further digital offers will become too bad.

References

1. Evstigneeva, G. A., Wegner, B.: O proekte sozdanija elektronnogo archiva russkich publikatsii po matematiki. Proceedings of LIBCOM 2002, Yershovo, November 2002.
2. Evstigneeva, G. A., Zemskov, A.: 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. Berlin: Springer. Lect. Notes Comput. Sci. 2730, 44–51 (2003).
3. Wegner, B.: EMANI — Leader and follower for the WDML, in New Developments in Electronic Publishing, Proceedings of the ECM 4 Satellite Conference on Electronic Publishing, Stockholm, June 25–27, 2004, 161–169, 2005.