

MathDiss International

| | |
|---|---|
| Background | 1 |
| Dissertations Online und its results | 3 |
| Specific features of mathematical dissertations | 3 |
| The concept of MathDiss International | 4 |
| Objectives and implementation of the project MathDiss International | 6 |
| International Cooperation | 7 |
| Conclusion | 8 |
| contact persons | 8 |

Background

The development of a scientific community depends by and large on its scientific productivity; the actuality of its research questions discussed within scientific communities and by the work being done or completed on these research questions, in particular in the mathematical and science fields, is decisive for the technological development of our communities for tomorrow. This is underlined by the fact that mathematics has in the meantime taken on the position of a key technology. Keeping pace with scientific research internationally is a challenge which has to be constantly strived for.

In this context dissertations¹ present a significant indicator for scientific research. These scientific papers document actual research questions in the laboratories and on the writing desks on the one hand, and on the other hand they reflect methods employed in research. Therefore all science communities must have a vital interest to *take in* this information and equally to *make it public* so that it can serve as the a basic product for innovation².

Information however entails the danger of being experienced by the individual as nothing else but like an uncontrolled flood. Thus concepts have to be developed enabling individual scientist to come to terms more adequately with this enormous flow of information. This appears to be more a characteristic experienced in our western society communities, whereas scientists from countries of the so-called “Third World” may be experiencing difficulties more with the opposite situation, namely not information getting through to them. In this context mass electronic information and communication (IuK) plays a decisive role. It can compensate the deficits at both ends: it can optimize precision selection of information on the one hand, and on the other hand

¹ The Deutsche Bibliothek (DDB) in the Federal Republic of Germany counts 20.000 copies pro annum

² To quote a program of the Federal German government 1996 -2000

it can with the greatest of ease open information sources which by traditional methods would only be possible – if at all - under great effort.

It therefore does not come as a surprise that particular academic communities for mathematics and the sciences in the Federal Republic of Germany have recognized the need to innovate in this field as well as to assess the possibilities. These communities founded the IuK-Initiative² to accompany and develop innovations in information und communication processes in academic disciplines - last but not least in the exact sciences to be able to accompany constructively and productively. An development example in the Deutsche Mathematiker Vereinigung (DMV)³ is *Math-Net*⁴ - an Internet information service for mathematicians, which serves as a reference source. Likewise we would like to point out that *PhysNet*⁵ - the Physics Department and Document Network serves the same purpose for the Deutsche Physikalische Gesellschaft (DPG)⁶.

Within this process scientists fulfill a double role: on the one hand they deliver information, and on the other hand and receive it as well. For the first time on the level of electronic communication the mediating role of the publishing companies and other producers of literature has been relativated, in particular because financing library aquisitions is becoming increasingly too expensive for universities in Germany and elsewhere. It is thus good advice to encourage scientists to take some of their interests into their own hands here. The Deutsche Mathematiker Vereinigung (DMV) in cooperation with the Österreichische Mathematiker Gesellschaft (ÖMG)⁷ has given these questions an independent subject-specific platform⁸ catering for the individual sections of the annual conferences.

Electronically archived scientific and general documents play a decisive role here. Especially the non-profit sector opens possibilities for development and design, whereby electronically represented doctorate dissertations are attributed a central, even an exemplary, role. For this purpose a new term has been established in the international context for the field of electronic papers: Electronic 'Theses and Dissertations' (ETDs). On the one hand ETDs are a suitable object for studying new workflows and problems related to electronic publication, and on the other hand one is also dealing with scientifically substantial contents in these documents. The annual ETD conferences of the Networked Digital Library of Theses and Dissertations (NDLTD)⁹ bear wit-

² <http://www.iuk-initiative.org>

³ http://www.mathematik.uni-bielefeld.de/DMV/mathnet_de.html

⁴ <http://www.Math-Net.org>

⁵ <http://physnet.uni-oldenburg.de/PhysNet/physnet.html>

⁶ <http://www.dpg-physik.de>

⁷ <http://www.mat.univie.ac.at/~oemg/>

⁸ <http://elib.zib.de/IuK-DMV/>

⁹ <http://www.ndltd.org>

ness that these chances have been recognized by the communities and have themselves become the object of various surveys and research projects.

Dissertations Online und its results

The recently completed DFG-project Dissertations Online¹⁰, initiated and developed by the DMV, has developed solutions for electronically archiving and publishing dissertations. The mentioned project focuses on

- making digital dissertations accessible via metadata
- adaption of a search engine towards further metadata types
- interdisciplinary networking of servers as search engines
- development of document type definitions (DTDs) and automatic generation of information for conversion to XML
- inclusion of multimedia elements (chemistry) into digital dissertations,
- setup of a user-oriented information system, and
- adaptation to present-day library techniques and procedures.

During the execution of the project in the years 1998-2000 it became clear that for the objective of reaching a unified solution for treating dissertations electronically there exist barriers across all disciplines, not even taking into consideration traditional criteria or procedures individual from disciplines for defining dissertations – which we consider to be a marginal but solvable aspect.

The central problem still is the aspect of a variety of differing data formats. The employed data formats are necessarily geared to presentation necessities for the scientific results in question. Insofar one cannot expect uniformity here in the middle term. Every doctorate will make use of a language appropriate to the level of complexity of the subject matter in question. The central issue of the above-mentioned project *Dissertationen Online* was documents compiled under *Microsoft Word*, a system designed for format requirements in particular in the humanities. Thus subsequent priority was given to working on documents produced in the LaTeX-format, the standard format for mathematical-scientific documents.

Specific features of mathematical dissertations

In the case of Word-based dissertations one can assume that there are no difficulties converting them into SMGL as archive format when they do not entail components of mathematical formulae¹¹. However, transformation of LaTeX-documents into SGML is principally possible but at

¹⁰ <http://www.dissonline.org>

¹¹ Every new Word version entails new surprises here, however!

present only to a limited degree. On the one hand one notes that e.g. formulae in relatively clear structured LaTeX degenerate to images and graphics. On the other hand however, a mathematical dissertation not seldom consists of about 90% formulae content, and thus it is clear that a transformation into SMGL would create a complexity that cannot be desirable or practicable.

This problem cannot be simply overlooked, as the actual statistics (latest stand September 2001) in the University and Polytechnics sector registered at the Deutschen Bibliothek (DDB)¹² present a clear picture. The quota of online dissertations in mathematics lies at almost 20%, and it is only surpassed by dissertations from chemistry (23%) and physics (22%). Altogether they constitute an interdisciplinary quota of about 10%.

This problematic condition is also being recognized more and more in international mathematics and science communities. The latest developments are taking this into account and are reducing the burden of publishing and making available mathematical and scientific contents in the World Wide Web (WWW), while at the same time preserving the semantic information of the LaTeX-codes, e.g. OpenMath¹³ or MathML (Mathematical Markup Language)¹⁴. The international congress held in Berkeley on the *Future of Mathematical Communication*¹⁵ demonstrated that these “languages” are being developed. As MathML however needs further semantic information not entailed in the LaTeX-code, the presence of a mathematically correct and automatic conversion method is at present illusionary. It will take a number of years before these languages have matured and reached the *acceptance* and the *distribution* (concerning science text compilation) already enjoyed by LaTeX in the science world today.

The concept of MathDiss International

The state of the formats and the archiving of an optimum of information *at present* has to meet the demands of long-time archiving, and for an unforeseeable time in the future they will have to also contain ASCII-LaTeX-files, which provisionally serve as source files and for presentation purposes (and not(!) SGML-derivates). Even in the case of LaTeX-Files a problematic variety of formats is noted due to text authors referring to their own self-developed input files. In all cases it is very necessary here to discuss homogenization. Nonetheless, transitional solutions are being worked out by experts in the United States for preserving authentic LaTeX-Texts, as the translation of such texts into subsequent market-adequate products is viewed as a problem which will eventually be solved without doubt, as already recognized by the International Mathematical Un-

¹² <http://www.ddb.de>

¹³ <http://www.nag.co.uk/projects/openmath/omsoc/>

¹⁴ <http://www.w3.org/TR/REC-MathML/>

¹⁵ <http://www.msri.org/activities/events/9900/fmc99/>

ion (IMU)¹⁶, the *Committee on Electronic Information and Communication* (CEIC)¹⁷ und the Los Alamos Servers¹⁸. This led to the development of a machine on which electronic mathematics and science papers are “conserved” in LaTeX format. Dissertations, however, only play a minor role on this server. Furthermore, the framework conditions of this American machine by no means meets appropriate demands in Germany concerning standards, (metadata, authentication etc.) agreed upon and established with *Dissertationen Online* and science libraries e.g. DDB und SUB Göttingen.

File format does not only have problematic consequences for long-time archiving but also plays a decisive role for automatic recherche of texts. As a full text search is not an adequate substitute for text recherche but instead more than not leads to a flood of information one can hardly cope with, the *Document Type Definitions* (DTDs) are of central importance alongside the original metadata here. From a pragmatic point of view one must imperative that such DTD information is automatically compiled. Consequently, it would be necessary to develop or adapt one’s own tools for each and every text-producing program (and possibly for every subsequent update).

In the project *Dissertationen Online*¹⁹ this has been completed with satisfaction for Word documents and has been achieved only with limited satisfaction for LaTeX documents, which was allocated secondary priority. A Word-parallel solution for LaTeX-Files was hereby looked for to fundamentally guarantee comparability with Word documents. Finally, it was not taken into consideration that LaTeX has its own programming possibilities for generating metainformation - however in the XML-affined form. It was not possible to ignore the fact that as a rule Word is employed in other academic fields, namely in the humanities as compared to TeX or LaTeX.

Once again it is apparent that a document server for a homogenous science discipline as well as for a homogenous format in this context allows more simple solutions when during production of the document (i.e. through specific indexing) appropriate preparations are made. It is conceivable that on the level of the LaTeX language glossaries, indexes, bibliographies etc. can be exported relatively easy into a database allowing exactly defined text recherche and search methods. As the demands in such homogenous science discipline worlds hardly differ on an international scale, one can expect a high degree of international cooperation here, which also has a synergetic side effect for the development of professional conversion tools.

¹⁶ <http://elib.zib.de/IMU/>

¹⁷ http://elib.zib.de/IMU/IMU_Committees/CEIC.html

¹⁸ <http://xxx.lanl.gov>

¹⁹ Neue Office-Versionen bereiten hierbei neue Probleme!

Building on the results of *Dissertationen Online*, these thoughts are fundamental for setting up an international mathematics dissertation server (MathDiss International)²⁰. This project sponsored by the Deutschen Forschungsgemeinschaft (DFG) is a joint venture of the Institut für Mathematik at the Gerhard-Mercator-Universität Duisburg and the Staats- und Universitätsbibliothek of the Bundesland Niedersachsen in Göttingen (SUB)²¹. The last mentioned will cater for the service and continue its development after this present project phase has been completed.

Objectives and implementation of the project MathDiss International

The objectives of the project can be divided into two areas.

▪ Search options

Subject-specific search in mathematical documents eventually underlies in the no change in time. Scientists do not employ other methods of literature *recherche* just because search medium or publication form has changed. In the light of the functionalities of the text set system LaTeX (assuming it is employed here), an adequate search appears feasible. Viewing their use as most natural, these functionalities generate automatically new files that can be referred to for adequate retrieval of automatic contents lists, for processing bibliographies with BibTeX, for compiling of the index with MakeIndex etc.. In particular we are dealing with the information of enhanced value relevance concerning the following search options:

- abstract,
- mathematical subject classification (MSC),
- contents,
- bibliographies,
- index and
- key words and glossary

One aspect however has to be also mentioned here. It is still necessary to „educate“ the authors i.e. the doctorates to write on a literary level of high quality in conjunction to their cognitive scientific activities. Already at the beginning of literary production the candidate must be aware of the circumstances and the conditions the dissertation will be later subjected to. In the case of electronic publication the prerequisites for high-quality subject-matter specific retrieval of sophisticated content have to be set as a basis. This fundamental requisite, that is source files enriched with information, leads us straight to the next central issue:

²⁰ <http://www.ub.uni-duisburg.de/mathdiss/>

²¹ <http://www.sub.uni-goettingen.de/>

- Archiving information

As already described in the general section of this paper, appropriate conversion of archive formats is not possible without difficulties. For this reason we have to preserve all information available. This is only possible by storing the source file of all the required styles and use package files.

One recognizes that participation in *MathDiss* International (as a library, an institution or as an individual) is linked to efforts which have to be worked out in the forefront. What are the advantages, however, which can be achieved by participation?

Dissertations passed on to this enterprise are stored in the linked Allegro-database, which is fed with the generated metadata. These are on the one hand the metadata, which meet the requirements of the Deutsche Bibliothek (DDB) for submitting electronic dissertations, and on the other hand the metadata are automatically generated in agreement to the NDLTD-Initiative²². Thus the dissertation can be found in both an international and an interdisciplinary data bank. This is of particular importance, because actual reports or comments inform us that there is the possibility to refer to electronic dissertations independent of time or place but the quality of retrieval suffers, be this due to unspecific search options (such as only author and title) or be this due to split searches onto different servers. Thanks to international links and the specialized focus of the services of *MathDiss* International this constitutes a problem that we can however come to terms with.

A further *bonbon* for doctorates can also be offered. Due to an agreement between the central publishing organ MATH²³ and the central publication of the European Mathematical Society (EMS)²⁴ all dissertations, which meet the standards and demands of the project *MathDiss* International and its services, are published and reviewed in the central periodical MATH. This also enables international recognition of the dissertations.

International Cooperation

A further working aspect of our project is entailed in the naming of the project. The objective of internationality is to be fulfilled. For this purpose it has been and still is our intention to build up international cooperation. Besides the already mentioned cooperation with NDLTD and thus also with the Open Archive Initiative (OAI)²⁵ further contacts have already been established with other institutions, in particular with other countries with German-speaking communities outside of Germany. To promote cooperation with Austria, appropriate steps have already been initiated.

²² <http://www.ndltd.org>

²³ <http://www.zblmath.fiz-karlsruhe.de/MATH/home>

²⁴ <http://www.emis.de/>

²⁵ <http://www.openarchives.org>

However, we are still waiting for an official decision of the Austrian Government Ministry on this matter. In Switzerland a competent and prominent partner, namely ETH Zurich, has been won for this project. Alongside further contacts with other northern and eastern European states we would also like to mention the cooperation with Cellule MathDoc from the French University of Grenoble, which publishes an annual list of mathematical dissertations in France²⁶ Further going cooperation has already been agreed upon with the University of Uppsala. This cooperation and further talks are to be upheld in the near future.

Conclusion

On the whole one can assert that with *MathDiss* International an instrument of particular importance for mathematics and international mathematical research has been developed and implemented. The user aspect stands in the foreground of this initiative. The central organization structure of this service enables one to quickly obtain an overview of actual research results. This project is however not only conceived as a research project specific to mathematics, but also as a model for building a document server homogenous in subject matter and format on an international level.

We invite all interested persons to critically participate in the development of this service and would like to offer all doctorates of mathematics wherever they may be in the world the opportunity of publishing their dissertations on *MathDiss* international server, also including automatic referencing in the central publishing organ MATH.

contact persons

Prof. Dr. Günter Törner
toerner@math.uni-duisburg.de

Dipl.-Math. Thorsten Bahne
bahne@math.uni-duisburg.de

Gerhard-Mercator-Universität Duisburg
Fakultät 4 – Naturwissenschaften; Institut für Mathematik

BD Hans-J. Becker
becker@mail.sub.uni-goettingen.de

Dr. Thomas Fischer
fischer@mail.sub.uni-goettingen.de

SUB Göttingen

URL von *MathDiss* International: <http://www.ub.uni-duisburg.de/mathdiss/>

²⁶ <http://www-mathdoc.ujf-grenoble.fr/these.html>