DOCUMENTA MATHEMATICA:
EFFICIENT PRODUCTION OF MATHEMATICAL LITERATURE

Report given on the Workshop

"The Future of Mathematical Communication: 1999", Berkeley [1]

Dec. 1 – 5, 1999

ULF REHMANN

DOCUMENTA MATHEMATICA is an *electronically produced* Journal for Mathematics, existing for almost four years. It has developed a production method which proved its capability and efficiency in 1998, when it was used to produce the Conference Proceedings of the International Congress of Mathematicians 1998 (Berlin) as an Extra Volume.

For the first time, the Proceedings of an ICM have been produced completely electronically. Using the facilities of DOCUMENTA MATHEMATICA, the contents of the scientific lectures were made available without charge on the internet, already before the congress started, at

http://www.mathematik.uni-bielefeld.de/documenta/

Two volumes of the printed version were distributed to the participants at the beginning of the congress, the volume containing material which had to be gathered during the congress was printed and distributed about three months after the congress.

This fact was considered

> *a remarkable demonstration of the potential to publish a major book at minimal cost without commercial assistance.*[2]

The Conference Proceedings of the ICM'98 consist of the

Vol. I:	"PLENARY LECTURES AND CEREMONIES"	662 pages,
Vol. II, III:	"INVITED LECTURES"	881 resp. 825 pages.

All volumes are in hardcover, 6 times 9 inches, containing many pictures including

[1] Some answers to questions asked at the end of the talk have been integrated here.

[2] in terms of the IMU president (at that time), David Mumford, in his speech at the closing ceremony of the ICM'98, cf. Proceedings of the ICM 1998, vol. I, p. 53.

32 colored pages.

The production of the proceedings was organized by myself, the formatting and compiling of the scientific part (2228 pages) as well was totally done by myself, without assistance by other persons. The non scientific part (140 pages) was edited in collaboration with Gerd Fischer and Hannes Stoppel.

For the purpose of printing I produced, for every volume, a single postscript file containing the ready-to-print data, which was posted on the Internet. [3]

A printing company in Rosenheim (Bavaria) then downloaded the files and manufactured from them the books, including printing, binding and shipping.

Volume I contains the manuscripts by the 21 invited "Plenary Speakers", approximately 20 pages each, the speeches given at the opening and closing ceremonies, the laudationes and the talks given by the Fields medalists and by the Rolf Nevanlinna price winner, as well as the list of participants and an appendix. Volumes II and III contain the manuscripts by the 170 invited "Section Speakers", approximately 10 pages each.

ORGANIZATION AND TIMING OF THE PRODUCTION:

NOV. 14, 97: Meeting with the printer to discuss the details of the production.
Time needed: 1.5 days trip: Bielefeld–Rosenheim.[4]

JAN. 98, FIRST WEEK: Construction of the TEX format files for the proceedings, their posting on the Internet, posting of an html form for manuscript submitting.
Time needed: 2 afternoons, approximately 8 hours.[5]

JULY 1, 98: Deadline for manuscript submission by the authors.

JULY 1–15, 98: MAIN WORK.
Formating of manuscripts, returning of the formated pages to the authors for proof reading and for last corrections, collecting the copyright forms from the authors, construction of the postscript files for volumes II, III including table of contents and index.
Time needed: 80 hours.

JULY 15, 98: Posting of the postscript files for volumes II, III for the printer.Time needed: marginal.

[3] In fact, for each volume there were two more postscript files for the book cover and back.

[4] It was important to make detailed plans that early; the printer was able to plan his work suitably and to buy the necessary paper for a good price.

[5] The format files and the Internet channel were built on the basis of the DOCUMENTA installation, which at that time was already in existence and operational for three years, therefore only little time was needed for this purpose.

July 18–23, 98:
Checking and correcting the proofs posting of the corrected files on the Internet for the printer, posting of the corrected total volumes II, III and of the "Plenary Lectures" (main part of volume I) on the Internet for the public.Time needed: 20 hours.

Starting July 26, 98: Final printing and binding of volumes II, III.

Starting August 13, 98: Shipping of volumes II, III from Rosenheim to Berlin.

August 18–27, 98: ICM'98, at the beginning: Distribution of volumes II, III to the participants.

Oct. 28–Nov. 13, 98: Formating of volume I, Internet posting of the final volume for the printer, final proofreading and corrections.
Time needed: approximately 40 hours.[6]

Starting Nov. 16, 98: Printing, binding and shipping of volume I by the printer.

Production of the ready to print postscript files

The construction of the postscript files used the concepts of Documenta Mathematica:

The authors had to submit their manuscripts via an interactive web form or by email. In order to enable them to pre-format their files, four style files were offered on the web in the TEX dialects TEX, Ams-Tex, Latex 2.09, Latex 2e.

All authors but one submitted TEX files using the offered methods, about 50 % of them were pre-formated using one of the Documenta Mathematica style files.

During the time from July 1. (deadline for authors) through July 15. (deadline given by the printer), approximately 170 manuscripts of 10 pages or more each by the "Invited Section Speakers" and 21 manuscripts of 20 pages or more each (in fact a total of 2200 pages) had to be formated, each article had to be returned to the author for galley proof reading, the revised versions had to replace the earlier version in the final volumes, the table of contents as well as the author index had to be constructed for volumes II, III, and, last but not least, the authors had to fill out and submit the copyright form – transferring the right of distribution of their manuscript to Documenta Mathematica.[7]

[6] Collecting the missing documents for volume I - everything except the "Plenary lectures", approximately 140 pages, required more time than expected. This work was done by G. Fischer and H. Stoppel, their time needed is not considered here.

[7] The process described here was slightly more complicated: Many manuscripts contained graphics, which were to be included via encapsulated postscript files or similar, for some manuscripts there have been 10 or more such files.

This was done essentially semi-automatically with some programs developed by myself for that purpose. The tools used (Emacs editor, Perl, C compiler, TEX, format converters, fonts, web server and browsers are all freely available on the Internet.[8]

1. With some Perl program, all those TEX files were re-formated, which were not already formated in the prescribed Documenta Mathematica style.

Then, from the editor ("Emacs") and by means of some previewer, the final formating was done: Checking of makeup, positioning of formulae, spell checking etc.

2. After fixing its format, the file was automatically[9] stored into the data basis for the final volumes, an email to the author containing the re-formated version was prepared, asking for galley proof reading and further details like abstract and classification and for submitting of the Documenta Mathematica copyright form. This email could be send out (possibly after adding some more specific requests) immediately after formating.

3. Two further Perl programs and one C program then constructed from the pre-formated data basis the table of content and the author index for the total volumes.[10] After that, from the data bases for the volumes the final postscript files were constructed using standard Unix programs like "make" and some techniques earlier developed for Documenta Mathematica[11]

4. Upon email receipt of modified files from the authors, changes were taken over semi-automatically[12] into the data basis. The time for each manuscript was significantly less than 5 minutes on average.

The Perl and Lisp programs mentioned above were developed by me during the procedure described above (approximately 4 hours inside of the 80 hours men-

[8] At production time, the commercial Digital Unix had been used as an operating system, but today (December 1999) all the tasks can be done with the required speed on an average Linux PC, whose operating system is freely available as well.

[9] by some Lisp program, as 'elisp' is the control language of Emacs

[10] For this it was not possible to use standard TEX tools, as those usually require the same TEX dialect for all documents. Documenta permits various dialects since otherwise we would loose the cooperation of the authors – every author strictly believes in his/her TEX dialect. It is much easier to develop tools like the above than to translate manuscripts from one TEX dialect into another.

[11] Automatic determination of the TEX dialect of a file, determination of the multiplicity of the necessary TEX translations in order to resolve all references, computation of the foot lines for each manuscript which contain information about the manuscript's respective first and last page, computation of the contents and index pages, collection of all the resulting dvi files and translation into postscript.

[12] Changes by the authors were displayed on two simultaneous windows - Emacs allows to take over or discard any changes by a keystroke, a mouse click walks from one difference to the next etc.

tioned), the C program (for the analysis of dvi files) I had developed earlier as a tool for handling the data of Documenta Mathematica.

Reasons for Finishing the Edition in Time

1. *The cooperation by the authors.* The authors were very cooperative. The submitted TEX files were almost all perfectly or even excellently prepared and required little work. (Only one manuscript had to be typed, since it was submitted on paper. Approximately 50 % of all authors did use the Documenta Mathematica style files.

2. *The usage of suitable style files.* The style files were slight modifications of the Documenta Mathematica style files and only adjusted to the format of the proceedings. The styles offered covered all the styles which were used by the authors, non had to be transcribed into another dialect. Apparently, all styles offered were necessary, since every dialect occurred sufficiently often: Plain TEX: 13 %, Amstex: 16 %, Latex 2e: 60 %, Latex 2.09: 11 %.

I assume if one of the styles would not have been offered or supported, then it would not have been possible to finish the production in time.

3. *The possibility and ability to automatize essential steps of the production.* This is closely related to 2. The Documenta Mathematica style files have been designed with respect to the following points of view:

i.) They are small, hence easily handled.[13] (This is also important for getting the cooperation of the authors.)

ii.) They are not restrictive, i.e., they allow the authors to use their own macros. This implies that they can easily be put on any file of its dialect in a semiautomatic procedure, for example by using e.g. an interactive input masque on the web or by using any graphical user interface.

iii.) They allow in an easy way an automatic further processing of the files equipped with them, e.g., in order to produce derived data like tables of content or author indices etc.

Conclusions

So far, we spoke about things which happened in the past. This is a workshop on the “Future of Mathematical Communication”. Talking about the future, the French writer André Malraux recommends:

[13] The average line number of the style files is around 150, while the average number of the style files of a well know commercial publisher is around 3500, and of another electronic journal: approximately 1100.

If you want to read the future, you should first spell the past.

Did we spell the past correctly? Concerning the facts mentioned, yes, I guess. However, we did not mention the pace of progress yet, which will allow to reduce and to simplify the work for journal editing even further in the future:

Earlier ICM proceedings have been published commercially, but rarely sooner than one year after the congress, sometimes two years later. The recent one, produced non commercially but using electronic automatization, did appear completely three months after the congress.

Hence the production time has decreased considerabely, at least by a factor of 4.

In fact, the technical production time has decreased by a much higher factor: two of the three volumes appeared already before the congress.

Concerning the time (80 hours) and the actual work I spent on the formating of the scientific part of the production, I have the strong impression that it is possible to automatize the whole process of technical formating (for TEX manuscripts) by designing and writing appropriate software.

Then the necessary time for technical book and journal production (besides writing and editing) would go down to almost zero: Only little work and time would be necessary to physically produce such things, and that work could be done by virtually every scientific editor – even with little or no technical experience – without professional help.

Production cost is low, production time is neglegible, if the production takes advantage of the benefits of electronic automatization, this holds for print and for electronic production.

The printing cost for the ICM'98 Proceedings was around

US $ 25.69 per three volume set

(ca. 2300 pages including many pictures and 32 color pages).[14]

That is, production of mathematical literature can be quick and affordable, if the methods of electronic publishing are used adequately.

[14] After adding the share of my salary including tax and social security and all other 'hidden costs', the production cost increases just by US $ 2.65 to a total of US $ 28.34. One of the questions at the end of the the talk was: "But the lights have been on in your office?" Taking this as a question concerning the infrastructure costs induced by my office at the university building: If one assumes monthly costs of US $ 1000.00 (probably too high), this adds production costs of less than US $ .30 per each triple volume set, since I worked approximatly during 1 month for the proceedings production. Of course, during that time, I fullfilled my obligations at my department as well.