

DOCUMENTATION ISSUES FOR MATHEMATICS IN THE DIGITAL AGE*

By Pierre Bérard

The impact of licensing science and technology digital resources on collection development, archiving and access for non-affiliated users are major issues for libraries. They were discussed at the 2002 open session of the Science & Technology Section of the International Federation of Library Associations in Glasgow. Being an end-user, I would like to address these documentation issues the other way around and ask:

How should the needs of end-users influence library policies, in particular concerning licensing, selecting material and archiving?

Mathematicians have always been very much involved in documentation issues. Academic institutions publish top mathematics journals (e.g. Acta Mathematica, Annals of Mathematics, Publications Mathématiques de l'Institut des Hautes Études Scientifiques), mathematical societies publish long-standing international journals, the American Mathematical Society and the European Mathematical Society are very much involved in databases, several committees discuss documentation issues (e.g. the Committee on Electronic Information and Communication—CEIC of the International Mathematical Union and the Electronic Publishing Committee of the European Mathematical Society), etc. In France, documentation issues in mathematics are handled in a collaborative way by the Réseau National des Bibliothèques de Mathématiques-RNBM¹ and by the Cellule MathDoc, an Institute on scientific information and communication (a joint institute, Centre National de la Recherche Scientifique-CNRS and Université Joseph Fourier, Grenoble).

I am aware that other scientific communities are very active on documentation issues as well, and I acknowledge the fact that we should share our experience and avoid acting in isolation. In this article, I shall concentrate on documentation issues for mathematics. I will discuss the importance of the scholarly literature for

* Paper presented at the 68th General Conference and Council of the International Federation of Library Associations and Institutions, August, 16 – 24, 2002 in Glasgow

¹ French network of mathematics libraries.

mathematics, the needs of mathematicians – as far as documentation is concerned – and I will present some of the actions that have been undertaken, in particular in France.

Note: The URLs of institutions or services mentioned in this paper are provided in the last section.

Mathematics and its scholarly literature

Mathematics is unique among the sciences in its dependence on the scholarly literature. Mathematics has indeed flourished along the past 2 500 years because key literature was passed from generation to generation. Time scales in mathematics and in other fields differ: ideas and techniques from decades or even one hundred years ago are relevant, sometimes crucial, for solving contemporary problems in mathematics or for applying mathematics to other sciences and technology.

Since the 17th century mathematics has played an important role in physical sciences and in technology by providing tools to organize, analyse, compute and predict. More recently, mathematics has assumed a similar role in quantitative aspects of the life sciences. The role of mathematics in technological development is increasing every day. Documentation issues in mathematics are therefore important not only for the discipline itself, but for the other sciences and for technology as well.

Let me recall a few figures. About 15 journals contained mathematical papers in 1700 and they were 200 by the end of the 18th century. Less than 1 000 mathematics papers were published every year in the second half of the 19th century. Today, the reviewing databases Mathematical Reviews and Zentralblatt MATH both publish about 75 000 items every year and they analyse nearly 600 mathematics journals from cover to cover, among thousands of sources (over 1 500 journals, as well as books, conference proceedings, etc).

The whole corpus of mathematical scholarly literature actually makes up the environment in which mathematicians live, in which they look for inspiration, for examples to mold their intuition, for tools to work out their ideas.

It is impossible today for mathematicians to have a global knowledge of their field. They need to move from one item to another in the scholarly literature, looking for concepts, ideas, examples, results and proofs. More importantly, mathematicians, as well as those who use or apply mathematics, need to rely on previously established results they may not always be able to check by themselves.

What do mathematicians need?

How should these needs influence library policies?

The needs of mathematicians –as far as documentation is concerned– follow from their dependence on scholarly literature. These needs should, I believe, directly influence institution policies concerning documentation.

Mathematicians need reliable sources

Because mathematics so much depends on its past literature, journals have played a major role since their emergence in the 18th century by providing validation and storage of key mathematical pieces. As far as the long-term welfare of mathematics is concerned, this is far more important than the social role [7] describes.

The robustness of the validation system has so far been guaranteed by the diversity of independent, and very often long-standing², reputable peer-reviewed journals.

In this framework, publishers compete for markets; journals, groups and individuals compete for recognition or ranking. This competition and the diversity of journals (for-profit, not-for-profit) have participated in keeping quality high, in improving services and in avoiding that groups or schools take undue control over the publication process. This competition has also, to some extent, contributed in controlling subscriptions fees (many top mathematics journals are published under the auspices of academic institutions, with reasonable subscription fees).

The digital age brings new opportunities –from archival eprint servers to purely electronic journals– to disseminate and access literature. At the same time, publication and validation are becoming increasingly detached.

Our main concern, for the long-term welfare of mathematics, should be to preserve the diversity of well-identified reliable sources.

In this perspective, consortial agreements and bundling practices have negative side effects. They indeed tend to bring more and more money to bigger publishers to the detriment of smaller ones (in particular academic publishers). They concur to reducing the diversity of the journal offer and to letting commercial considerations overcome scientific considerations when selecting sources. As stated in *Best current practices: Recommendations on Electronic Information and Communication*, a set of recommendations issued by the Committee on Electronic

² Among the oldest mathematics journals, let me mention for example, the Journal für die reine und angewandte Mathematik (Crelle's journal) created in 1826 and the Journal de mathématiques pures et appliquées (Liouville's journal) created in 1836.

Information and Communication (International Mathematical Union), “When institutions are forced to accept or reject large collections of scholarly literature covering many different disciplines, the decisions are less likely to be made by scholars.” ([1] Item 15). Consortial agreements may also concur to distorting the publishing landscape, as explained in [7], Chapter 10. As a matter of fact, the transition from selling/buying print collections to licensing electronic access may in the long-run endanger access to reliable sources.

The Cellule MathDoc and the Réseau National des Bibliothèques de Mathématiques apply and promote the following principles (which are, I believe, very much in the spirit of SPARC).

- Consortial agreements should be based on lists of journals rather than based on whole catalogue offers (the “Big Deals” described by K. Frazier [5]). These lists should be established according to scientific and relevance criteria.
- Enough print subscriptions should be maintained as long as long-term access to digital content is not secured.
- As far as documentation is concerned, institutions should take global budget approaches leaving room for smaller publishers, thus concurring to maintaining the diversity of journal offers. They should also keep a reasonable balance between journals and books collections.
- Institutions should, as far as possible, support the actions undertaken by editorial committees against excessive subscription. Libraries should inform the academic communities on pricing issues and institutions should encourage scientists to take journal policies into account before deciding to submit papers or to do referee work.
- Scientists and specialised librarians should be invited to participate in the negotiations with publishers.

Scientific communication has been made much easier by the possibility to exchange digital files over the internet, and in particular by the emergence of eprint servers. As a by-product, the number of multi-version papers tends to increase. To provide version identification and well identified validation procedures for electronic only documents are compulsory to insure reliability of electronic communication.

Mathematicians need to have access to the scholarly literature over a very wide time-span

In this respect, the digital era brings both concerns and hopes; concerns because preservation, archiving and long-term access issues are becoming even more intricate and hopes because we may dream of a large virtual library on the internet.

Concerns

- Because mathematics so heavily depends on its past literature, preservation and archiving are of prime importance. Libraries (maybe not individually but on a global scale) have up to now played a central role by providing preservation and archiving of print material. New technical problems arise in the digital age. In view of the present licensing practices, our main concern today is however of a political nature.

Who will in the future be responsible for preserving and archiving the mathematical heritage?

- Libraries have not only cared for preservation and archiving. They have provided long-term access for their patrons and, more generally, for the general public. Publishers used not to take care of providing long-term access. With the digital era, this is changing.

Will the mathematical production remain part of our common heritage or will it be gradually confiscated by private interests?

In view of the importance of mathematical scholarly literature for the discipline itself, and for the advancement of science and technology as well, it is mandatory that the tasks of preserving, archiving and providing long-term access to the mathematical heritage be shared in an international endeavour by academic communities and institutions (using open standards). These tasks cannot be left to the sole responsibility of commercial groups whose main concerns are revenues, not documentation issues. The International Mathematical Union has endorsed a recommendation in this direction, [1] Recommendation 14.

Interesting experiences to work on these important issues are underway, for example Project EUCLID (Cornell University) on the academic side and the EMANI project, an international joint venture between libraries (Cornell, Göttingen, Orsay, Tsinghua) and publishers in the Springer Verlag group.

Hopes

- The mathematical community pursues the project of creating a *Digital Mathematics Library* of the scholarly literature. This endeavour has been endorsed by the International Mathematical Union and by mathematical societies world-wide.

The Digital Mathematics Library will be a comprehensive collection comprising both retro-digitised and natively digital documents. It is meant to serve scientific communities, students and, more generally, citizens world-wide, by providing easy and efficient access (in particular cross-linking) to the mathematical heritage. The

Digital Mathematics Library is expected to have a strong impact on the way mathematics is done and used in the 21st century, an impact that may turn out to be as important as the emergence of journals three centuries ago.

The realization of this project will consist of three phases (design, implementation and sustained operation). Questions to be addressed concern scientific choices, technical choices, copyright issues, archiving and sustained operation (including future migrations of formats).

Meetings have already been held to prepare for the project: San Diego (USA) in January 2002, Berlingen (Switzerland) in April 2002 and Washington (USA) in July 2002. Several funding agencies have shown their interest in the project, including the Centre National de la Recherche Scientifique–CNRS (France), the Deutsche Forschungsgemeinschaft–DFG (Germany) and the National Science Foundation–NSF (USA).

The general idea is that the raw data (raw images of the scanned documents) will be available at low cost to all on the internet and that mathematics societies, institutions or publishers will provide added value services and enhancements (cross-linking, comments) possibly on a paying basis (for a possible scenario, see [3]). It is expected that the sustaining effort will be undertaken by academic institutions and societies world-wide. The project will require co-operations between academic institutions, libraries and publishers. I believe that we must be very careful that the retro-digitised mathematical heritage be not confiscated by private interests.

The French mathematical community is participating in this endeavour with the NUMDAM³ programme, on which I will briefly report later on.

- Libraries have in the past played an important role by providing free access not only to their patrons but also to a wider audience (walk-in users). It seems desirable to extend this to the internet in some way.

The International Mathematical Union endorsed a recommendation for *Unrestricted access* to data such as Table of contents, Abstracts, Keywords (and Bibliographies whenever possible) and a recommendation for *Eventual free access* (free access to the full-texts after a suitable time-span), see [1] Recommendations 11 and 12. As a matter of fact some journals are already committed to this goal.

³ NUMérisation de Documents Anciens Mathématiques (Digitisation of Ancient Mathematics Documents).

Mathematicians need organization and linking

- The need for tools to organize and approach a growing scholarly literature was soon recognized by mathematicians (and others). This led to the emergence of classification schemes and databases. Here are some of them (see [6] for a more general study and [9] for more information on the Répertoire Bibliographique des Sciences Mathématiques).
 - Jahrbuch über die Fortschritte der Mathematik (1868–1942), established by Carl Ohrtmann and Felix Müller.
 - Répertoire Bibliographique des Sciences Mathématiques (1894–1912), established by the Société mathématique de France and chaired by Henri Poincaré.
 - Zentralblatt für Mathematik und ihre Grenzgebiete⁴ (1931 →), established by Otto Neugebauer.
 - Mathematical reviews (1940 →), founded by the American Mathematical Society.
 - Mathematics section, Referativnyi Zhurnal (1952 →).

The pioneers clearly expressed their motivations.

Das Ziel, das uns vorschwebte, war einerseits: Demjenigen, der nicht in der Lage ist, alle auf dem umfangreichen Gebiet der Mathematik vorkommenden Erscheinungen selbstständig zu verfolgen, ein Mittel zu geben, sich wenigstens einen allgemeinen Überblick über das Fortschreiten der Wissenschaft zu verschaffen. Andererseits: dem gelehrten Forscher seine Arbeit bei Auffindung des bereits Bekannten zu erleichtern⁵.

Carl Ohrtmann and Felix Müller

Jahrbuch über die Fortschritte der Mathematik

Toute classification est une théorie déguisée, et ce n'est pourtant qu'en classant les faits qu'on pourra se mouvoir dans le dédale sans s'égarer. Ceux qui méconnaîtront cette vérité ne marcheront qu'à tâtons, revenant sans cesse sur leurs pas, refaisant cent fois le même chemin ...⁶

Henri Poincaré

Le libre examen en matière scientifique (1909)

⁴ It is today published under the title Zentralblatt MATH by the European Mathematical Society, the Fachinformationszentrum–Karlsruhe and Springer Verlag.

⁵ “Our intention was on the one hand: To provide a tool for those, who are not able to follow all publications on the comprehensive field of mathematics, and to gain a general overview about the development of the science. On the other hand: It should help the active scientist to find out known facts.”

⁶ “Any classification is a disguised theory. It is however only by classifying facts that one will be able to move in the maze without getting lost. Those who will not recognize this truth will only grope along, coming back onto their own steps, taking a hundred times the same path ...”

Today, both indexing and reviewing services – Mathematical Reviews and Zentralblatt MATH – cover all of mathematical scholarly literature. They collaborate on a common classification scheme (Mathematics Subject Classification 2000) which is used by the whole mathematical community. They benefit from the work of thousands of reviewers world-wide. Both services offer external links from reviewed items to full-text. By including the bibliographies of papers, together with internal links, in the corresponding reviews they will eventually provide a tool to explore the interconnections within the literature (see [8] for the state of the art at Mathematical Reviews) thus providing alternatives to the large databanks of commercial publishers.

Indexing and reviewing databases therefore represent a strategic resource for mathematics. This is why I believe that the existence of two competing databases of high quality is beneficial for the mathematical community world-wide (improved quality, better access conditions). As a matter of fact, these ideas have been endorsed by the European Mathematical Society and others. In France, the Cellule MathDoc takes an active part in the transformation of Zentralblatt MATH into a large European research infrastructure for the benefit of mathematics world-wide, under the auspices of the European Mathematical Society.

- Cross-linking is a major issue for publishing in the digital age. This is an expensive process and one should be very careful that cross-linking services do not become the monopoly of the bigger publishers (for an interesting scenario, see [4]). We expect that Mathematical Reviews and Zentralblatt MATH will both play an important role in providing links to digital documents, in particular in the Digital Mathematics Library project⁷.

From principles to practice

The Cellule MathDoc, in collaboration with the Réseau National des Bibliothèques de Mathématiques, has set up services for the French mathematics community. I will illustrate our approach with three examples which are directly related to the issues addressed here (for more details on our activities, see [2]).

Databases: Mathematical Reviews and Zentralblatt MATH

- As already mentioned, the Cellule MathDoc has since 1996 taken a very active part in the transformation of Zentralblatt MATH into a large European research

⁷ Note that the Jahrbuch is being digitised and already provides over 12 000 links to digitised papers in print collections.

infrastructure, for the benefit of mathematics world-wide, under the auspices of the European Mathematical Society.

In this framework, we have negotiated a general consortial agreement with the Zentralblatt MATH database which goes beyond mere licensing. French institutions subscribe within the general agreement which in particular allows small centres to benefit from reduced subscriptions fees (still with a site licence). The Cellule MathDoc has set up a network of three national mirrors for Zentralblatt MATH to ensure robustness of access (controlled by IP numbers). The search and display engine is provided by the Cellule MathDoc (as for the international mirrors) and the database is updated every month. This mirror system guarantees long term availability to the data. The network of national mirrors also provides a better integration of resources (journals in the Zentralblatt MATH database are linked to the common catalogue of mathematics serials which provides localization and hence makes document delivery easier).

- A general consortial agreement – of the usual licensing kind – has been passed between the Réseau National des Bibliothèques de Mathématiques and the Mathematical Reviews database.

Current contents service and electronic journals

- The Cellule MathDoc negotiated a national consortial agreement with a data provider. Table of contents are downloaded every week by FTP to our server and indexed locally. The system is built on in-house tools and offers the following functionalities:
 - Table of contents of about 450 core mathematics journals and 450 journals in related fields.
 - Search and display interfaces (browsing by title, search by author, keyword).
 - Alert service (email address of the user / list of serials).
 - Information concerning the serials (localization in French libraries through the common catalogue of mathematics serials).
 - Hyperlinks to web pages of journals when available, links to the “OpenResolver” service.

Access to the current contents service is controlled by IP numbers (under the responsibility of the Cellule MathDoc) and is open to the whole mathematics community including occasional users. As in the case of Zentralblatt MATH, this agreement goes beyond mere licensing.

- A national consortial agreement has been passed by the Réseau National des Bibliothèques de Mathématiques for the electronic access to a subset of mathematics journals in LINK (Springer Verlag group). The agreement is based on a set of print subscriptions in libraries belonging to the network; electronic access is granted to all mathematics institutes in France (from the work desk of the mathematicians, control by IP numbers); access for external users; the extra cost is covered by a special grant from the Centre National de la Recherche Scientifique–CNRS. Negotiations with other publishers are in progress.

The next step would be to globally reconsider the number of print subscriptions for each individual journal before negotiating.

Let me add that one of the main libraries in the Réseau National des Bibliothèques de Mathématiques (Bibliothèque Jacques Hadamard, Orsay) is a partner of the Electronic Mathematics Archiving Network Initiative – EMANI project and hence involved in the difficult issues of the digital age: archiving, preservation and long-term access for digital content.

Digitisation programme

The general goal of the NUMDAM⁸ programme is to digitise the mathematics scholarly literature published in France in print form and to participate in the Digital Mathematics Library international project mentioned previously. The first phase of the programme concerns five internationally known serials. Subsequent phases will concern other journals and important seminar series.

The following technical choices take the needs of end-users into account:

- Pages scanned at 600 dpi to fully recover complex formulas and to allow later treatments depending on technological progress (such as structure or formulas recognition).
- Character recognition of text to allow searches.
- Segmentation to allow precise access to articles.
- Bibliographies of articles are key-boarded and tagged. They are included in the database of articles.
- Links to the mathematics databases will be provided whenever possible (at the level of digitised articles and at the level of bibliographic items contained in the digitised articles).

⁸ NUMérisation de Documents Anciens Mathématiques / Digitisation of Ancient Mathematics Documents.

We promote the following principles:

- The meta data (cataloguing data, abstracts, bibliographies, hidden OCREd full-text) of the digitised documents will be freely accessible to all on the internet.
- The full-text (image mode) will be freely available on the internet after a suitable moving-wall set up to insure the financial stability of journals.

The recommended moving-wall is 5 years (articles published in year N to be freely available from N+5 onwards). The digital collections are scheduled to be incremented with natively digital production (with the same moving-wall). These principles are implemented on a case-by-case basis, after suitable negotiation with the institutions and publishers involved. Copyright issues are also addressed. One difficulty we have encountered with commercial publishers concerns cross-linking from bibliographic items (to be compared with [7], Chapter 11)

Discussions with mathematicians and historians of mathematics are underway to enhance the digitised collections with comments.

Conclusion

For us, producers and users of contents, documentation is clearly a scientific issue. Due in particular to international competition, it is also a political and an economic issue. The concentration phenomenon in the publishing business shows that documentation has become a commercial issue. We cannot minimise the fact that publishing activities – academic or commercial - and development of services with added value (databases, search facilities, cross-linking) require know-how, they cost money and generate money and jobs.

The welfare of mathematics, as a scientific discipline and as a service provider for other sciences and technology, will, in the future, very much depend on how well we continue to enrich, to preserve, and to archive the mathematical scholarly literature, our common heritage. The welfare of mathematics will also depend on the fact that the mathematical scholarly literature remains part of humanity's knowledge commons and that it is not gradually confiscated for the benefit of restricted circles.

We must be very careful that the choices and decisions we make today do not endanger the mathematical heritage in the future.

URL's⁹

American Mathematical Society (AMS)

<http://www.ams.org/>

Cellule MathDoc

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

Centre National de la Recherche Scientifique (CNRS)

<http://www.cnrs.fr/>

Committee on Electronic Information and Communication (CEIC)

International Mathematical Union

<http://www.ceic.math.ca/>

Deutsche Forschungsgemeinschaft (DFG)

<http://www.dfg.de/english/index.html>

Electronic Mathematics Archiving Network Initiative (EMANI)

<http://www.springer.de/press/companynews/emani.html>

European Mathematical Society (EMS)

<http://www.emis.de/>

International Mathematical Union (IMU)

<http://www.mathunion.org/>

Mathematical Reviews

<http://www.ams.org/mathscinet>

National Science Foundation (NSF)

<http://www.nsf.gov/>

NUMérisation de Documents Anciens Mathématiques (NUMDAM)

<http://www-mathdoc.ujf-grenoble.fr/NUMDAM/>

Project EUCLID

<http://projecteuclid.org/>

Réseau National des Bibliothèques de Mathématiques (RNBM)

<http://www.biblio.math.jussieu.fr/reseau.html>

Société de Mathématiques Appliquées et Industrielles (SMAI)

<http://smai.emath.fr/>

⁹ Last visited, May 2002.

Société Mathématique de France (SMF)

<http://smf.emath.fr/>

The Scholarly Publishing and Academic Resources Coalition (SPARC)

<http://www.arl.org/sparc/>

Université Joseph Fourier (UJF)

<http://www.ujf-grenoble.fr/>

Zentralblatt MATH

<http://www-irma.u-strasbg.fr/ZMATH/>

References

[1] CEIC –. Best Current Practices: Recommendations on Electronic Information and Communication (2002), endorsed by the Executive Committee of the International Mathematical Union on April 13, 2002 (CEIC, Committee on Electronic Information and Communication)

http://www.ceic.math.ca/ceic_docs/best_practices/Best-Practices.pdf (last visited May 2002)

[2] Cellule MathDoc –. 2001 Annual Report

<http://www-mathdoc.ujf-grenoble.fr/Activites/MDC-Rapport2001-public.pdf> (last visited May 2002)

[3] Ewing, John –. Twenty centuries of mathematics: _igitising and disseminating the past mathematical literature

<http://www.ams.org/ewing> (last visited May 2002)

[4] Ewing, John –. In defence of caution, Journal of the Association of Learned and Professional Society Publishers (3/28/2002)

<http://www.ams.org/ewing> (last visited May 2002)

[5] Frazier, Kenneth –. The librarian's dilemma : contemplating the costs of the Big deal, D-Lib Magazine, 7 (2001)

<http://www.dlib.org/dlib/march01/frazier/03frazier.html> (last visited October 2001)

[6] Goldstein, Catherine –. Sur quelques pratiques de l'information mathématique, *Philosophia Scientiæ* 5 (2001), 125-160

[7] Guédon, Jean-Claude –. In Oldenburg's Long Shadow: Librarians, Research Scientists, Publishers, and the Control on Scientific Publishing, ARL Proceedings 138

<http://www.arl.org/arl/proceedings/138/guedon.html> (last visited May 2002)

[8] Kister, Jane –. Reference lists and citations in the Mathematical Reviews database, Notices of the AMS (American Mathematical Society), October 2001, page 965

<http://www.ams.org/notices/200109/commentary.pdf> (last visited May 2002)

[9] Rollet, Laurent – Nabonnand, Philippe –. Une bibliographie mathématique idéale ? Le répertoire bibliographique des sciences mathématiques, Gazette des mathématiciens, Soc. Math. France 92 (2002), 11–25

Pierre Bérard

Institut Fourier umr 5582 ujf-cnrs & Cellule MathDoc
ums 5638 ujf-cnrs

Université Joseph Fourier – BP 53

F – 38041 Grenoble Cedex 9

France

Pierre.Berard@ujf-grenoble.fr

<http://www-fourier.ujf-grenoble.fr/~pberard/>