

The World Digital Mathematics Library

Thierry Bouche

Institut Fourier & Cellule Mathdoc
Université Joseph-Fourier (Grenoble)

ICM 2014
Seoul, 2014-08-20

A citation of Jean-Pierre Serre (according to Michel Broué)

« *Les mathématiciens se contentent de mettre leur production à la disposition de tous, comme sur des étagères où l'on peut venir se servir.* »

“Mathematicians just make their results available to everyone as if they were on shelves, waiting to be fetched.”

A remark from a colleague to whom I showed a preview of the first online posting of Numdam : *Why isn't it all there ?*

Numdam - Numérisation de documents anciens mathématiques

Recherche et téléchargement d'archives de revues mathématiques numérisées

Revue de l'Enseignement Mathématique (1981-1982)

Revue de l'Enseignement Mathématique (1981-1982)

2003

Numdam - Recherche et téléchargement d'archives de revues mathématiques numérisées

Revue de l'Enseignement Mathématique (1981-1982)

Revue de l'Enseignement Mathématique (1981-1982)

Revue de l'Enseignement Mathématique (1981-1982)

2014

The mathematical corpus

- Mathematical *validated* literature never becomes obsolete
(Old results are not superseded by newer ones: they are their foundation, full proofs are sometimes never written twice!)
- It's valid only as a *whole*, building a wide network of references
- It's useful to other sciences in an *asynchronous* fashion

The mathematical corpus is the set of all (potentially) referenceable published works

- ⇒ It must be carefully archived, indexed and preserved
- ⇒ It must be widely accessible over the long term

The reference mathematical library

We thus need a *reference library*, which should be

- comprehensive
- up-to-date
- well organized
- long lasting
- widely open
- easy to use for non-mathematicians

And digital, with power tools opening new paths for research & serendipity !

EuDML, an outline

- European innovation project, programme CIP ICT PSP focused on “Open access to scientific information”
- Duration 3 years; Budget 1.6 M€; Work 40.5 years
- Closed on January 2013
- Follow-up (unfunded) : EuDML initiative, an international association formed under the auspices of the EMS

The screenshot shows the EuDML website interface. At the top, there is a search bar with the text "Title, Author, Keyword, Citation, Date" and a search button. Below the search bar are navigation tabs: Home, Advanced Search, Browse by Subject, Browse by Journals, and Refs Lookup. The "Advanced Search" tab is selected. A search box contains the text "Enter your search terms to get started" and a search button. Below the search box are options for "Advanced Search" and "Embed EuDML Search on Your Website". A sidebar on the right contains a "Search Tips" section with bullet points: "search is case and diacritics insensitive", "search is performed on exact words as long (maximum 4 characters)", "browse by keywords with quick cluster", "Information Theorem", "Mathematics Browser", "uniformization AMC browser", "advanced", and "You can also search in arXiv". Below the search box, it states "EuDML is currently indexing 233856 items across 14 collections". At the bottom, there are three sections: "What is EuDML?", "Features" (listing search and explore, find related items, and save and share), and "Recent Notes" (listing a note about a research group on algebraic points and page processing).

About the Project | Partners | Developer API | Feedback | Version 2.0

EuDML

The screenshot shows a search result for a paper by Jean-Pierre Serre. The title is "Quelques applications du théorème de densité de Chebotarev". The author is "Jean-Pierre Serre". The publication is "Publications Mathématiques de l'IHÉS (1961)", volume 54, page 123-201, with ISSN 0073-8301. There are buttons for "Access Full Article" and "Open in full text". Below the title, there is a "Cite" section with a citation in MLA format: "Serre, Jean-Pierre. 'Quelques applications du théorème de densité de Chebotarev.' Publications Mathématiques de l'IHÉS 54 (1961): 123-201. <http://euclid.amsjournals.org/10.1090/S...>". There is also a "References" section with a list of references. On the right side, there are sections for "Paper Details", "Add to Personal Lists", "First Similar Documents", and "Subjects". The "Subjects" section lists "Abelian varieties and schemes" and "Curves".

Eudml results

In a nutshell, we produced

- A critical mass in content
233 000 texts of all supported types \approx 6% of math. corpus
- A cooperation network
Now founding members of the EuDML initiative members
- A math-savvy fully functional digital library
MathML metadata, math mining, MSC, links to/from math databases
- A good looking Web site with unique navigation tools
adapted to our user community
Internal and external deep interlinking, MSC browsing, reference lookup
- A number of productivity and interoperability devices
enabling the main service
Some production ready, some more experimental
- EuDML initiative organizational model and policies
Under the strong control of science through EMS

EuDML policy

A content provider needs to agree on these

- 1 EuDML content is scientifically validated and formally published in final form
- 2 The digital content is physically hosted at one of the partner institutions (local DML)
- 3 It is accessible after a reasonable moving wall (which typically ranges from zero, or Open Access, to 5 years)

Lessons learned

- I fully second the NRC report that a dedicated core team located at one institution would be much more efficient
- The basic layer is well understood and works when content partners obey a minimal set of technical requirements.
This means that scaling from 6% to 30% of the mathematical corpus is at hand
The most pressing demand from mathematicians is to scale to 1
- But still many resources won't be easily available either because
 - they are not digitised
 - they are not professionally digitised
 - item-level metadata is lacking
 - there is no way to harvest item-level metadata
 - the content owner/host is not willing to provide (or was not asked)

All these are workable, but need quite an amount of effort (technical, legal, political) and support from the community (cf. my call to all journal editors in November 2014 issue of the *Notices of the AMS*)

Principles to keep in mind

- The archive should be secured against most turbulences (partners changing commercial strategy, single State agency or private foundation funding, etc.)
⇒ should be a distributed and replicated *physical* archive
- Correctness matters to math : it is of high importance to let intelligent agents generate derived mathematical knowledge that would be machine readable for enhancing service (spontaneous crowd-sourcing, OCR, structure/semantics recognition. . .) but this process should be transparent to the users and never hide the original, unmodified sources
- We are talking of an infrastructure for research that will be the daily working tool of mathematicians worldwide. What it really needs is long-term institutional support just like your university or department library today !