

NUMDAM

*and other
digitisation activities
at
MathDoc*

Thierry BOUCHE
Université Joseph Fourier (Grenoble)

WDML workshop
Grenoble
March 11th 2005

NUMDAM

I. Presentation

1. Definition
2. Principles
3. Collections
4. Main features

II. MathDoc: other related projects

1. LiNum, RBSM & Liouville
2. The Gallica frontend
3. The mini-DML project

1. DEFINITION

*Digitise for archiving and delivery the backrun
of academic mathematical journals*

Archiving : integral scan from the first page up to the last one of each volume, including covers, plates, unbound leafs, ads, etc.

Delivery : one multipage file per article, access through tables of contents browsing or searching, freely downloadable after a variable moving wall depending on the publisher's taste.

2. PRINCIPLES

- Integrity :
 - full backrun, no editorial choice;
 - scan of every page at high resolution (600 dpi) black & white (text), grey or colour when applicable ;
 - page format reproduced.
- Interactivity :
 - detailed structured metadata captured, allowing to search over basic bibliographical datas plus full (literary) text and cited references.
 - a hyperlink network places the article in proper context: bibliographies, reviews, errata, etc.

3. COLLECTIONS – CURRENTLY ONLINE

Title	Since	Owner	Volumes	Pages	Articles
<i>Ann. inst. Fourier</i>	1949	Assoc. A.I.F.	156	51 054	1 811
<i>Ann. I.H.P.</i> [†]	1930-1964	I.H.P.	71	5 860	147
<i>Ann. math. Blaise-Pascal</i>	1994-2002	Labo/UBP	19	2 554	165
<i>Ann. Sci. École norm. sup.</i>	1864	É.N.S.*	295	68 898	1 867
<i>Ann. univ. Grenoble</i> [†]	1945-1948	UJF	3	1 006	47
<i>Bull. Soc. math. France</i>	1872	S.M.F.	167	45 774	2 608
<i>Mém. Soc. math. France</i>	1964	S.M.F.	134	18 118	396
<i>Journées É.D.P.</i>	1974	C.N.R.S.	31	5 976	514
<i>Publ. math. I.H.É.S.</i>	1959	I.H.É.S.*	92	17 424	344
* Contract with a commercial publisher.				216 664	7 899

3. COLLECTIONS – FORTHCOMING

Title	Since	Owner	Volumes	Pages	Articles
<i>Ann. Fac. Sci. Toulouse</i>	1887-2000	U.P.S.	207	36 052	1 035
Sém. Prob. Strasbourg	1967-2002	Labo IRMA*	37	17 352	1 254
Séminaires I.H.P. [†]	1953-1985	??	146	20 000	1 800
Sém. Bourbaki	1948-2002	Assoc. N. B.*	44	17 000	893
<i>Ann. Gergonne</i>	1810-1831	D.P.	22	8 000	935
<i>Ann. I.H.P. sér. A</i>	1964-2000	I.H.P.*	64	25 000	1 125
<i>Ann. I.H.P. sér. B</i>	1964-2000	I.H.P.*	39	20 000	936
<i>Ann. I.H.P. sér. C</i>	1985-2000	I.H.P.*	16	10 000	368
<i>Ann. Sc. Norm. Pisa</i>	1871-2001	SNS	88		
<i>Rev. Stat. appl.</i>	1953-2000	SFS	52	21 890	1 080
<i>Ann. Fac. sci. Univ. Clermont</i>	1962-1993	UBP	37		
<i>Compositio Math.</i>	1935-1997	Fund. C.	70	40 000	
...					

3. NUMDAM FUTURE?

- Older, multidisciplinary journals?
- Books, Ph.D. thesis & other monographs?
- Manuscripts, rare items (Bourbaki archives...)?
- Collaborations with other partners (Regional/European funding, more journals from other countries)?
- Real time integration of current metadata from live journals?

4. MAIN FEATURES

- Access to the articles through browsing or searching.
- “Full notice” compliant with CEIC best practice (full bibliographic reference + abstract + bibliography + links freely available).
- Full text available for more than 90% of the collection (after journal-dependent moving wall).
- Download unit: full articles. Download formats (indirect DjVu & linearised PDF) allow for page-by-page downloading.
- Dual nature of the interface: HTML metadata with links vs. multipage faithful image of the full text as download unit.

4. MAIN FEATURES (LINKS)

- As many usefull links as possible:
 - Same author, same journal volume.
 - Full text when available.
 - Reviewing service when matched (MR, ZM, JFM, SPS, ...)
 - Errata/original article when necessary.
 - From a cited item to its reviews (MR, ZM, JFM) or to its full notice (NUMDAM, many more in a near future thanks to the mini-DML).
 - To (NUMDAM) articles that cite the given article.

For 7 899 articles, we have 4 970 MR ids, 5 388 ZM ids and 2 031 JFM ids.

4 877 of them have a formalised reference list, amounting for 78 413 cited items. 79% have a ZM id, 66% an MR id, while only 2% have a JFM id. We have less than 100 erratum relations, but more than 5 600 direct NUMDAM links.

4. MAIN FEATURES (MOST RECENT)

- Enhanced search engine (returns linked page numbers for words in full text; option to search expressions present on the same page).
- Full catalogue available through OAI-PMH server.
- Browsing interface “cloaked” for Web crawlers for better indexing where scarce metadata is available.
- Links to NUMDAM from MathSciNet, Zentralblatt, RBSM, Google, Yahoo, OAI agregators.

II. MATHDOC : OTHER RELATED PROJECTS

- **LiNum : Livres numérisés mathématiques.**
2 577 freely accessible books, 651 digitised but copyrighted, provided by large digitisation centers : Gallica (Paris), *Digital Math Books Collection* (Cornell), *Historical Math Collection* (Ann Arbor), *Mathematica* (Göttingen), *Biblioteka Wirtualna Matematyki* (Warsaw), etc.
- **Le *Répertoire bibliographique des sciences mathématiques* (1894-1912).**
Collaboration Gallica, Paris (scan of the cards), laboratoire de philosophie et d'histoire des sciences, Nancy (structured keyboarding of the cards) and MathDoc, Grenoble (database, indexing, online interface).
- **The *Journal de mathématiques pures et appliquées*, aka *Journal de Liouville* (1836-1932) :** detailed cataloguing and indexation of the volumes digitised by Gallica.
- A list of similar resources is (badly) maintained on :
<http://www.numdam.org/ressnum.php>.

2. GALLICA FRONTEND

The BNF's server Gallica has a huge amount of valuable mathematics that are somewhat hidden by weak metadata policy.

Cellule MathDoc is building a user frontend so that Gallica's resources will be mini-DML compliant:

- Standalone browsing/searching interface for "opaque" works such as journals (JMPA, CRAS, BSM) and collected works of important mathematicians.
- One full notice per item (reprinted or original article) so that third parties (including our mini-DML) can link them.

3. THE MINI-DML PROJECT

- mini-DML : unified indexation of articles available in digital format, taking advantage on the general dissemination of XML/OAI-PMH technology.

With special emphasis on long-run journals for which a large amount of material never made it into recent review databases (JFM, ZM, MR) : *Annales de l'ENS* (1864 : NUMDAM), *Bulletin de la SMF* (1872 : NUMDAM), *Journal de Liouville* (1836 : Gallica/Elsevier), *Comptes rendus de l'Académie des sciences* (1835 : Gallica/Elsevier), *Monatshefte für Mathematik und Physik* (1890 : DIEPER@Graz/??/Springer), *Annals of Mathematics* (1884 : JSTOR/arXiv), *Philosophical Transactions of the Royal Society of London* (1776 : JSTOR), *Duke mathematical journal* (1935 : project Euclid), *Commentarii mathematici Helvetici* (1929 : DGZ), *Mathematische Annalen* (1869 : DGZ/Springer), *Mathematische Zeitschrift* (1918 : DGZ/Springer), *Journal de l'École polytechnique* (1795 : X), *Crelle's journal* (1826: GDZ), etc. But also digitally available texts that are seldom explicitly referenced (reprints—possibly in collected works, e.g. from Gallica, preprints—arXiv, current issues, . . .)

- Will there be one day a comprehensive DML database which could resolve all “near” matches for any cited reference string from any published paper?

Thank you!