

Math academic publishing efforts from France (NUMDAM, CEDRAM, etc.)

Thierry Bouche

Cellule MathDoc & institut Fourier, Grenoble (France)

CMDE 2006, Aveiro, August 2006

MathDoc's support for math publishers & researchers

I. NUMDAM

- 1 Presentation
- 2 More details
- 3 Open questions

III. "Registries"

- 7 Gallica-Math

II. CEDRAM

- 4 Presentation
- 5 Methods
- 6 Open questions
- 8 Outer digitised material
- 9 The mini-DML project

I. NUMDAM

- 1 Presentation
 - Definition
 - Overview
 - Issues
 - Context
 - Collections
- 2 More details
 - Principles
 - Copyright model
 - Collections
 - NUMDAM future?
- 3 Open questions

Definition

Digitise for archiving and delivery the backrun of academic mathematical journals

Archiving : Integral scanning 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.

Definition

Digitise for archiving and delivery the backrun of academic mathematical journals

Archiving : Integral scanning 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.

Goals

- Keep a public-curated safe archive of the “french” mathematical cultural heritage.
- Enhance its visibility and accessibility.
- Provide an effective tool for current (and future) research.
- Support academic and independant publishing.

Specific issues

- **Mathematical literature never becomes obsolete.**
- It's useful to other sciences in an *asynchronous* fashion.
- It's valid only when considered as a *whole*, building a network of (international) references.
- Isolated academic journals starve against commercial concentration.
- The digital format makes it possible to keep the archive alive, with easy access and hyperlinks.

Specific issues

- **Mathematical literature never becomes obsolete.**
- **It's useful to other sciences in an *asynchronous* fashion.**
- It's valid only when considered as a *whole*, building a network of (international) references.
- Isolated academic journals starve against commercial concentration.
- The digital format makes it possible to keep the archive alive, with easy access and hyperlinks.

Specific issues

- Mathematical literature never becomes obsolete.
- It's useful to other sciences in an *asynchronous* fashion.
- It's valid only when considered as a *whole*, building a network of (international) references.
- Isolated academic journals starve against commercial concentration.
- The digital format makes it possible to keep the archive alive, with easy access and hyperlinks.

Specific issues

- Mathematical literature never becomes obsolete.
- It's useful to other sciences in an *asynchronous* fashion.
- It's valid only when considered as a *whole*, building a network of (international) references.
- Isolated academic journals starve against commercial concentration.
- The digital format makes it possible to keep the archive alive, with easy access and hyperlinks.

Specific issues

- Mathematical literature never becomes obsolete.
- It's useful to other sciences in an *asynchronous* fashion.
- It's valid only when considered as a *whole*, building a network of (international) references.
- Isolated academic journals starve against commercial concentration.
- The digital format makes it possible to keep the archive alive, with easy access and hyperlinks.

Context

- NUMDAM was initiated six years ago by MathDoc's directors: Pierre Bérard and Laurent Guillopé.
- Run by *Cellule de coordination documentaire nationale pour les mathématiques* (MathDoc), small service unit CNRS-UJF located in Grenoble. 4 part-time engineers, 2 advising mathematicians. Actual production outsourced.
- Only public funds so far.
- *Not a library service*: MathDoc owns no documents itself. NUMDAM is meant as a service to other parties, under the auspices of the French math community.
- *Not a new edition*.

Context

- NUMDAM was initiated six years ago by MathDoc's directors: Pierre Bérard and Laurent Guillopé.
- Run by *Cellule de coordination documentaire nationale pour les mathématiques* (MathDoc), small service unit CNRS-UJF located in Grenoble. 4 part-time engineers, 2 advising mathematicians. Actual production outsourced.
- Only public funds so far.
- *Not* a library service: MathDoc owns no documents itself. NUMDAM is meant as a service to other parties, under the auspices of the French math community.
- *Not a new edition.*

Context

- NUMDAM was initiated six years ago by MathDoc's directors: Pierre Bérard and Laurent Guillopé.
- Run by *Cellule de coordination documentaire nationale pour les mathématiques* (MathDoc), small service unit CNRS-UJF located in Grenoble. 4 part-time engineers, 2 advising mathematicians. Actual production outsourced.
- Only public funds so far.
- *Not* a library service: MathDoc owns no documents itself. NUMDAM is meant as a service to other parties, under the auspices of the French math community.
- *Not* a new edition.

Collections

- **First online posting fall 2003.**
- As of today: 14 serials, 1 series of proceedings, 22 important seminars (Strasbourg: Probability; Paris: Chevalley, Cartan, Pisot-Poitou...).
The first ever math journal (Gergonne, 1810-1831).
One European journal (Compositio Math.).
18 262 articles, 411 000 pages.
- Yet to come: 2 French journals of applied maths, , 2 European journals (Ann. Pisa, Rend. Padova), Bourbaki seminar...
More than 22 000 articles, 500 000 pages at the end of 2007.

Collections

- First online posting fall 2003.
- As of today: 14 serials, 1 series of proceedings, 22 important seminars (Strasbourg: Probability; Paris: Chevalley, Cartan, Pisot-Poitou...).
- The first ever math journal (Gergonne, 1810-1831).
- One European journal (Compositio Math.).
- 18 262 articles, 411 000 pages.
- Yet to come: 2 French journals of applied maths, 2 European journals (Ann. Pisa, Rend. Padova), Bourbaki seminar...
More than 22 000 articles, 500 000 pages at the end of 2007.

Collections

- First online posting fall 2003.
- As of today: 14 serials, 1 series of proceedings, 22 important seminars (Strasbourg: Probability; Paris: Chevalley, Cartan, Pisot-Poitou...).
The first ever math journal (Gergonne, 1810-1831).
One European journal (Compositio Math.).
18 262 articles, 411 000 pages.
- Yet to come: 2 French journals of applied maths, , 2 European journals (Ann. Pisa, Rend. Padova), Bourbaki seminar...
More than 22 000 articles, 500 000 pages at the end of 2007.

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.
- Unicode metadata.
- Clean separation between faithful articles' image and added metadata (PDF/DjVu article vs. navigation HTML stuff).

Principles – Interactivity

- Detailed structured metadata captured:
basic bibliographical data plus full (literary) text
and cited references.
- A hyperlink network places the article in proper context:
bibliographies, reviews, errata, etc.

Copyright model

- Electronic version under the title's owner control.
- Authors are asked to give their (exclusive) electronic copyright to them if still alive (preferred: academic institution).
- We contract with the copyright owner to allow access at www.numdam.org.
- Clear identification of the originating journal (logo, links, first page).
- Moving wall agreed by both parties (mean: 5, min: 0, max: 10); free access after the moving wall as a counterpart to public funding.

Currently online 1

Title	Period	Owner	Volumes	Pages	Articles
<i>Ann. Fac. Sci. Toulouse</i>	1887-2000	U.P.S.	207	36 052	1 035
<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. I.H.P. sér. A, B, C</i>	1964-2000	I.H.P.*	119	55 000	2 400
<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

* Contract with a commercial publisher.

.../...

Currently online 2

Title	Period	Owner	Volumes	Pages	Articles
<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
Sém. Proba Strasbourg	1967-2002	Labo IRMA*	37	17 352	1 254
Séminaires Paris [†]	1953-1985	??	146	37 700	2 647
<i>Ann. Gergonne</i> [†]	1810-1831	P.D.	22	8 700	1 083
<i>Compositio Math.</i>	1935-1996	Fund. C.	275	39 500	1 934
				411 000	18 262

* Contract with a commercial publisher.

Forthcoming

Title	Period	Owner	Volumes	Pages	Articles
Sém. Bourbaki	1948-2000	Assoc. N. B.*	44	17 000	893
<i>Ann. Scuola Norm. Sup. Pisa</i>	1871-2001	SNS	253	46 000	1 731
<i>Rev. Stat. appl.</i>	1953-2000	SFS	207	22 150	1 676
<i>Math. et Sciences humaines</i>	1962-2000	EHESS	148	15 000	1 000
<i>Ann. Fac. sci. Univ. Clermont</i>	1962-1993	UBP	37	5 000	
<i>Rend. Sem. Mat. Padova</i>	1930-2000	U. Pad.	115	25 000	
...					

Future developments

- **Books, Ph.D. thesis & other monographs.**
- Older, multidisciplinary journals?
- 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?

Future developments

- Books, Ph.D. thesis & other monographs.
- Older, multidisciplinary journals?
- 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?

Future developments

- Books, Ph.D. thesis & other monographs.
- Older, multidisciplinary journals?
- 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?

Future developments

- Books, Ph.D. thesis & other monographs.
- Older, multidisciplinary journals?
- 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?

Future developments

- Books, Ph.D. thesis & other monographs.
- Older, multidisciplinary journals?
- 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?

Open questions –Search engine

- How to manage user input text encodings, transliterations, and the like?
 - **Шафаревич** = Šafarevič = Safarevic ≠ Shafarevich?!
- Does there exist a format for OCRed maths that would allow to make the digitised content searchable?

Open questions –Search engine

- How to manage user input text encodings, transliterations, and the like?
 - **Шафаревич** = Šafarevič = Safarevic ≠ Shafarevich?!
- Does there exist a format for OCRed maths that would allow to make the digitised content searchable?
 - with a simple user syntax,
 - yet powerful enough to allow some kind of wildcards (integral of a function of a special class rather than $\int_{\Omega} f(x)$).

Open questions –General

- How to obtain a unified treatment of maths in titles, abstracts, bibliographies, and plain text?
- How to enhance access for (non-)mathematicians?
(Formula searching, vocabulary matching, structured content indexing. . .)
- Should we dig into the running text for references?
(formalised bibliographies are seldom present since the 1930s, standard since WWII only. . .)
- Do we prefer no link rather than few fuzzy ones?
- Will there be one day a comprehensive DML database which could resolve all “near” matches for any cited reference string from any published paper?

II. CEDRAM

- 4 Presentation
 - Definition
 - Goals
 - Context
 - Journals

- 5 Methods
 - Methods
 - Examples

- 6 Open questions

Definition

cedram

Centre de diffusion de revues académiques de mathématiques
“Center for diffusion of academic mathematical journals”

Support Tools for better production and management of independent or society research math journals.

Host High quality NUMDAM-like but journal-wide servers for the electronic edition.

Modularity Adapt to journal's needs.
Provide possibly an alternative to outsourcing to a commercial publisher.

Definition

cedram

Centre de diffusion de revues académiques de mathématiques

Support Tools for better production and management of independant or society research math journals.

Host High quality NUMDAM-like but journal-wide servers for the electronic edition.

Modularity Adapt to journal's needs.
Provide possibly an alternative to outsourcing to a commercial publisher.

Goals

Similar to NUMDAM, but tailored for born electronic publishing.

- Keep a public-curated safe archive of the “french” mathematical production.
- Enhance its visibility and accessibility.
- Provide an effective tool for current (and future) research.
- Support academic and independant publishing.

Goals

Similar to NUMDAM, but tailored for born electronic publishing.

- Keep a public-curated safe archive of the “french” mathematical production.
- Enhance its visibility and accessibility.
- Provide an effective tool for current (and future) research.
- Support academic and independant publishing.

Context

- Run by Cellule MathDoc staff, in cooperation with the editorial committees of the journals.
- Each journal remains completely independant (scientific content as well as financial balance).
- CEDRAM acts as a technical advisor and provider in the publishing process.
- *No* financial stream so far (but that might change).

Journals

Already online:

- *Annales de la faculté des sciences de Toulouse, mathématiques.*
- *Annales de l'institut Fourier* (Grenoble).
- *Annales mathématiques Blaise Pascal* (Clermont-Ferrand).

Soon online:

- *Journal de théorie des nombres de Bordeaux.*

Journals

Already online:

- *Annales de la faculté des sciences de Toulouse, mathématiques.*
- *Annales de l'institut Fourier* (Grenoble).
- *Annales mathématiques Blaise Pascal* (Clermont-Ferrand).

Soon online:

- *Journal de théorie des nombres de Bordeaux.*

Methods

Produce and manage the electronic edition of math journals.

- RUCHE: software for managing the refereeing process and author database.
- cedram.cls : a \LaTeX + \BibTeX driven production environment that automates whatever can be (page numbers, metadata generation, ...).
- Script-driven electronic edition from \LaTeX + \BibTeX sources.
- The paper edition is a by-product of the whole process.

Metadata from L^AT_EX source

Today (Latex2html)

Tomorrow? (tralics)

Open questions

- How to obtain a unified treatment of maths in titles, abstracts, bibliographies? (is MathML mature enough?)
- What is most important in bibliographies: links? cited author search?

Open questions

- How to obtain a unified treatment of maths in titles, abstracts, bibliographies? (is MathML mature enough?)
- What is most important in bibliographies: links? cited author search?

III. Registries

- 7 Gallica-Math

- 8 Outer digitised material

- 9 The mini-DML project
 - Abstract
 - Benefits
 - Shortcomings

Gallica-Math

The BNF's server Gallica has a huge amount of valuable mathematics that are somewhat hidden by weak metadata policy. Cellule MathDoc has built a user frontend so that Gallica's resources be mini-DML compliant and easily browsable:

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

Outer digitised material

- **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).

Outer digitised material

- 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 mini-DML project

The 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)...
- But also digitally available texts that are seldom explicitly referenced (reprints—possibly in collected works, e.g. from Gallica, preprints—arXiv, current issues, ...)



The mini-DML project

The 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)...
- But also digitally available texts that are seldom explicitly referenced (reprints—possibly in collected works, e.g. from Gallica, preprints—arXiv, current issues, ...)



The mini-DML project

The 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)...
- But also digitally available texts that are seldom explicitly referenced (reprints—possibly in collected works, e.g. from Gallica, preprints—arXiv, current issues, ...)

The mini-DML benefits

The mini-DML:

- allows interactive bibliographical search, returns a link to the full HTML notice of the articles (sort order : author-title),
- allows matching against its records: this will add many direct links to the NUMDAM/CEDRAM bibliographies (starting autumn 2006).

The mini-DML benefits

The mini-DML:

- allows interactive bibliographical search, returns a link to the full HTML notice of the articles (sort order : author-title),
- allows matching against its records: this will add many direct links to the NUMDAM/CEDRAM bibliographies (starting autumn 2006).

The mini-DML shortcomings

The mini-DML:

- does not provide author identification or variants matching (arXiv preprint with published paper, e.g.),
- might be fooled by text or math encoding variations,
- is very far from being exhaustive: there are large valuable repositories that don't make their data easily retrievable, please change your minds!

The mini-DML shortcomings

The mini-DML:

- does not provide author identification or variants matching (arXiv preprint with published paper, e.g.),
- might be fooled by text or math encoding variations,
- is very far from being exhaustive: there are large valuable repositories that don't make their data easily retrievable, please change your minds!



The mini-DML shortcomings

The mini-DML:

- does not provide author identification or variants matching (arXiv preprint with published paper, e.g.),
- might be fooled by text or math encoding variations,
- is very far from being exhaustive: there are large valuable repositories that don't make their data easily retrievable, please change your minds!

iThanks!