

Introducing the mini-DML project

Thierry Bouche

▶ To cite this version:

Thierry Bouche. Introducing the mini-DML project. ECM4 Satellite Conference EMANI/DML, Jun 2004, Stockholm, Sweden. 11 p.; ISBN 3-88127-107-4. hal-00347692

HAL Id: hal-00347692

https://hal.science/hal-00347692

Submitted on 16 Dec 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Introducing the

mini-DML

project

Thierry BOUCHE Université Joseph Fourier (Grenoble)

> WDML workshop Stockholm

June 27th 2004

Introduction

At the Göttingen meeting of the Digital mathematical library project (DML), in May 2004, the issue was raised that discovery and seamless access to the available digitised litterature was still a task to be acomplished. The ambitious project of a comprehensive registry of all ongoing digitisation activities in the field of mathematical research litterature was agreed upon, as well as the further investigation of many linking options to ease user's life. However, given the scope of those projects, their benefits can't be expected too soon.

Between the hope of a comprehensive DML with many efficient entry points and the actual dissemination of heterogeneous partial lists of available material, there is a path towards multiple distributed databases allowing integrated search, metadata exchange and powerful interlinking. The mini-DML project, initiated by Cellule MathDoc, is a prototypal implementation of this vision.

We will briefly discuss its roots, definition, and outline a proposal of good practice in metadata exposure and retrieval, as a mean to turn it (and its expected companion competitors) into a useful service.

1 Context

Scientific publishing has a long history:

- Scientific journals exist since 1665 (Journal des sçavans, Philosophical transactions).
- Mathematical journals exist since 1810 (*Annales* de Gergonne).
- Many digitisation projects are heavily working in order to provide that intellectual treasury to the digital age.
- Freely downloadable grey litterature is omnipresent.
- We face an exponentially growing amount of available material while
 - most researchers aren't even aware of this, and...
 - those who know have to bookmark hundreds of websites: repositories, specialised search engines, home pages, individual journal pages, and...
 - they need to figure a specific strategy to localise an item in each of these sites!

2 Catalogues & lists

Facing a similar situation, the mathematicians of the end of 19th century choosed to build lists of important published resources.

- Valentin's famous lost catalogue (1885-1910).
- The *Répertoire bibliographique des sciences mathématiques* (1894-1912, covering 19th century).
- The Jahrbuch über die Fortschritte der Mathematik (1868-1944).

And later:

- Zentralblatt-Math (1931-).
- Mathematical reviews (1940-).

Except in the (conjectural) first case, these lists are non-comprehensive and require an important intellectual input (choice, classification, review). Converted into databases, they are an invaluable utility for the working mathematician.

3 Where is the DML?

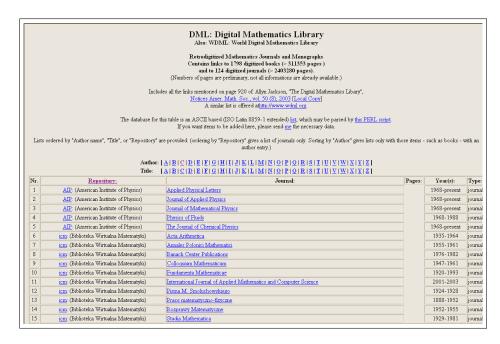
Let me risk the following definition.

The DML is the virtual library of all mathematical texts available digitally, both retrodigitised and born digital.

The difficulty which is addressed here is how to easily locate an item in the already vast DML, without being an expert in information technology.

How does an average mathematician know that such an important resource as *An-nals of Mathematics* is currently available

- at JSTOR from the first issue until 1994,
- at arXiv for certain years,
- at the journal home page (rare links to arXiv),
- at project Euclid since 2001.



 $\label{eq:figure 1-Ulf Rehmann's list of digitised mathematics. A mathematician's view of the DML.$

to here in UCC. This explains a possibly			rarious digitised journals, and also journals that we either currently or recently sub-
	ay whether or not digitised (and po	ssibly free) versions are available. In the ca	ere (beware - sometimes only the tables of contents and/or abstracts are available e of AMS journals (an institute-run suite of publications) and a few others the of
Journal	On-line articles from	Archived/digitised versions	
Acta Appl Math	46 (1997)		==1
Acta Math	144 (1980)		
Adv in Math	97 (1993)		
Aequationes Math	55 (1998)	GDZ (1968-1997)	_
Amer J Math	118 (1996)	JSTOR (1878-1995)	_
Amer Math Monthly		JSTOR (1894-1998)	
Ann Henri Poincaré	1 (2000)		
Ann Inst Fourier (Grenoble)	1 (1949) (free to now-5)	NUMDAM (1949-2000)	
Ann Math	140 (1994) (some free)	JSTOR (1884-1998)	
Ann Probab	24 (1996)	JSTOR (1973-2000)	
Ann Sci École Norm Sup	30 (1997)	NUMDAM (1864-2000)	
Bull Amer Math Soc	33 (1996) (free)		
Bull London Math Soc	29 (1997)		
Bull Sci Math	122 (1998)		
Bull Soc Math France	117 (1989)	NUMDAM (1872-2000)	
Cah Topol Géom Différ Catég	38 (1997)		
Canad J Math	49 (1997) (49 and 50 free		
Comment Math Helv	72 (1997)	GDZ (1929-1996)	
C R Acad Sci Paris Ser I Math	324 (1997)		
Chaos Solitons Fractals	1 (1991)		
Comm Math Phys	183 (no.3) (1997)		
Doc Math	1 (1996) (free)		
Electron Comm Prob	1 (1996) (free)		
Electron J Prob	1 (1996) (free)		
Ergodic Theory Dynam Systems	17 (1997)		
Funct Anal Appl	34 (2000)		
Geom Funct Anal	7 (1997)	GDZ (1991-1996)	

Figure 2 – A carefully arranged list of existing sources for mathematical serials (both born digital and retrodigitised). A librarian view of the DML.

How does an average mathematician know that the *Journal de théorie de nombres de Bordeaux* is currently available

- at Göttingen GDZ for the pre-series Séminaire...
- at EMIS for recent years (1994-),
- at the journal home page for recent years,
- · except two gaps.

It is currently almost impossible for one working researcher to keep some insight on the ongoing activities related to digitisation or digital publication in his field. It is common sense that collecting links to available resources is today's metaphore for our legacy library shelves. But library shelves are at *one* location and arranged in a coherent manner throughout the whole library. This is not the case with websites which all have their own way of arranging, naming, and access strategies to the material they archive.

4 The mini-DML project

The mini-DML project is currently in a very preliminary phase. A summary of the goals and techniques of the project follows.

Synopsis Collate in *one* place basic bibliographical data for any kind of mathematical digital article and make them accessible to the users through simple search or metadata retrieval. Where...

- Basic bibliographical data means
 - Author;
 - Title:
 - Date of publication;
 - Full bibliographical reference;
 - Unique identifier;
 - URL (preferably calculated from the above).
- Mathematical digital article means
 - Journal article;
 - Proceedings, seminar article;
 - Reprint, collected works;
 - Preprint;
 - Memoirs, books, slides... (not yet)
- Accessible to the users means
 - Simple unified search inteface;
 - Easy retrieval of the used metadata for further processing or different exposure to the public.

Collections A proof-of-concept implementation is presented here, based on a variety of sources of mathematical texts. The main emphasis is on long-run journals whose early production is widely unknown to *Math. Reviews, Zentralblatt-MATH* or even the *Jahrbuch*, with special interest towards current production and preprints on the other end.

- Annals of Mathematics (JSTOR URL guessed);
- NUMDAM journals (currently: six French titles, more to be added soon);
- One Gallica journal : *Journal de mathématiques pures et appliquées* (a.k.a. *Liouville*, up to 1880) ;
- Few Gallica complete works (Abel, Cauchy, Dirichlet, Fourier, Jacobi, Klein, Lagrange, Laguerre, Laplace, Möbius, Riemann);
- Some project Euclid journals (Duke...);
- Part of arXiv.

Method Retrieval of basic metadata through OAI protocol

- From our home-grown catalogues (NUMDAM, Gallica, Works);
- From OAI servers (Euclid, arXiv);
- Possibly from less structured sources (HTML tables of contents, OCRed tables of contents...).

Indexing with *Edbm-2* (MathDoc indexer).

It is forseen to eventually expose the remastered metadata through an OAI server using at least 3 schemes (Dublin core, minidml, proposed MR/ZM when stable). Of course, this will depend on the metadata provider's policy in this respect.

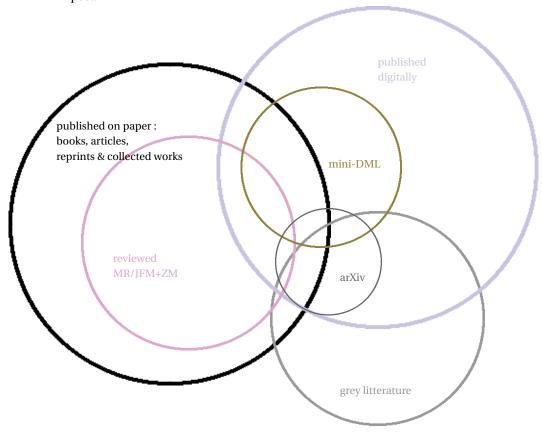


FIGURE 3 – A picture of how the various classes of mathematical documents overlap. The three main classes are: legacy paper, DML (available digitally) and grey litterature. The mini-DML scope has been zoomed in order to be visible.

Collaborations The size of the project makes it mandatory that the gathering of metadata be automatised. New material must be continuously added as indexed sites are updated. The best available solution is to find an agreement about a basic metadata scheme to be exchanged between content providers and indexers. We cannot expect that manual updates, and person-to-person exchanges permit to maintain such a service in the long term.

If you have scanned a long-run periodical, please provide

- a unique identifier for each article,
- structured basic metadata,
- a simple stable URL pointing towards an introductory HTML page for each article without restriction access;

and set up an OAI server to share your effort!

5 Where are Cauchy seminal papers?

Augustin Cauchy published a lot. But too early to be referenced in the mainstream databases, where only the publication of volumes of its collected works are acknowledged. Digging into the table of contents of the volumes allow to localise the actual quotation in a bibliography at NUMDAM website, e.g. The screen capture is of http://www.numdam.org/minidml/ as of september 2004.



 $\label{thm:conditional} Figure~4-Some~papers~from~Augustin~Cauchy's~Œuvres.$



FIGURE 5 – Once a paper is referenced in the mini-DML, it is possible to have a direct link from anywhere...

6 A case study

Louis de Branges de Bourcia claims he has at last completed a proof of Riemann hypothesis in 2004. This is for him an occasion to write an $Apology^1$ depicting his own

 $^{1. \} http://www.math.purdue.edu/ftp_pub/branges/apology.pdf\\$

path towards the conjecture. The paper contains 41 "bibliographical" references going back to 1730. I spent one full afternoon hunting for these references on the Internet. I present here the results, which show the variety of locations and strategies needed to eventually put your hands on the references.

Here is the list of cited items, grouped by entry type in citation order. How many strategies should try the working mathematician in order to know whether an item is digitally available, and to locate it when possible? The green links correspond to what you could expect to see one day in the mini-DML, i. e. papers already or soon available at large digitisation centres pertaining to the DML effort. The blue links also pertain to the DML, but as native digital documents (in our case : new editions of legacy papers, such as Hilbert's problems). The red links show the full URL. The other links are unexpected digitised items, to be found at some personnal home page or such, typically found thanks to Google!

One can dream of sort of fuzzy logic search engine that could be fed with any unstructured bibliographical citation and give back a list of near matches available in the DML. Two typical applications would be: add links to the full text of quoted items in the bibliographies of articles (batch mode); copy and paste a reference from an electronic document and look for it in the DML (interactive mode). This is something we will try to implement soon in the mini-DML.

- Thomas Stieltjes, "Recherches sur les fractions continues," Annales de la Faculté Scientifique de Toulouse 8 (1894), 1-122,
- and 9 (1895), 1-47.

⇒ NUMDAM II

A selection of such problems was presented by David Hilbert to the International Congress of Mathematicians which was held in 1900 in Paris: "Mathematical Problems," Bulletin of the American Mathematical Society 8 (1902), 437-479.

```
⇒ http://www.rinconmatematico.com/hilbert/hilbertproblems.doc
⇒ http://www.ams.org/bul1/2000-37-04/S0273-0979-00-00881-8/S0273-0979-00-00881
```

Jean-Pierre Gram, "Note sur les zéros de la fonction de Riemann," Acta Mathematica 27 (1903), 289-

⇒ http://www.actamathematica.org?not yet!

• Nikolai Sonine, "Recherches sur les fonctions cylindriques et le développement des fonctions continues en séries," Mathematische Annalen 16 (1880), 1-80.

- => http://134.76.163.65/servlet/digbib?template=view.html&id=26157&startpage=9&endpage=88&image-path=http://134.76.176.141/cobin/letgifsfly.cgi&image-subpath=/1342&image-subpath=1342&pagenumber=9&imageset-id=1 • by the author, "Self-reciprocal functions," Journal of Mathematical Analysis and Applications 9 (1964),
- Virginia Rovnyak in her thesis, "Self-reciprocal functions," Duke Mathematical Journal 33 (1966), 363-

⇒ http://projecteuclid.org/Dienst/UI/1.0/Summarize/euclid.dmj/1077376391

series of publications [by Hilbert], "Grundzüge einer allgemeinen Theorie der Integralgleichungen," Göttinger Nachrichten I (1904), 49-91,

//134.76.163.65/servlet/digbib?template=view.html&id=57552&startpage=55&endpage=97&image-path=http://134.76.161.141/cgi-bin/letgifsfly.cgi&image-subpath=/2133&image-subpath=2133&pagenumber=55&imageset-id=2133

• II (1904), 213-259,

378.

=> http://134.76.163.65/servlet/digbib?template=view.html&id=5757l&startpage=228&endpage=275&image-path=http: //134.76.176.141/cgi-bin/letgifsfly.cgi&image-subpath=/2133&image-subpath=2133&pagenumber=228&imageset-id=2133

III (1905), 307-338,

, http://134.76.163.65/servlet/digbib?template=view.html&id=62905&startpage=336&endpage=367&image-path=http //134.76.176.141/cgi-bin/letgifsfly.cgi&image-subpath=/2251&image-subpath=2251&pagenumber=336&imageset-id=2251apagenumber=33

IV (1906), 157-222,

= http://134.76.163.65/servlet/digbib?template=view.html&id=58133&startpage=169&endpage=240&image-path=http: //134.76.176.141/cgi-bin/letgifsfly.cgi&image-subpath=/2156&image-subpath=2156&pagenumber=169&imageset-id=2156

and V (1906), 439,480.

=> http://l34.76.163.65/servlet/digbib?template=view.html&id=58156&startpage=459&endpage=498&image=path=http://l34.76.176.141/cgi-bin/letgifsfly.cgi&image-subpath=/2156&image-subpath=2156&pagenumber=459&imageset-id=2156

• Erhard Schmidt, "Über die Anzahl der Primzahlen unter gegebener Grenze," Mathematische Annalen 57 (1903), 195-204.

http://134.76.163.65/servlet/digbib?template=view.html&id=36506&startpage=206&endpage=215&image-path //134.76.176.141/cgi-bin/letgifsfly.cgi&image-subpath=/1560&image-subpath=1560&pagenumber=206&imageset-id=1560

- The construction of Hilbert spaces of entire functions associated with Dirichlet zeta functions appeared as "Modular spaces of entire functions," Journal of Mathematical Analysis and Applications 44 (1973), 192-205.
- Hans Maass, "Über eine neue Art von nichtanalytischen automorphen Funktionen und die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen," Mathematische Annalen 121 (1949),

⇒ http://134.76.163.65/servlet/digbib?template=view.html&id=37241&startpage=145&endpage=187&image-path=http

- David Trutt and the author, "Orthogonal Newton polynomials," Advances in Mathematics 37 (1980),
- David Trutt and the author, "Meixner and Pollaczek spaces of entire functions", Journal of Mathematical Analysis and Applications 22 (1968), 12-24.
- Arne Beurling and Paul Malliavin at an International Symposium on Functional Analysis held at Stanford University. Their results "On Fourier transforms of measures with compact support" appear in Acta Mathematica 107 (1962), 291-392.

The publication of Newton's Principia Mathematica in 1687

=> http://gallica.bnf.fr/document?0=N00336

 $\Rightarrow Newton\ online = \ http://dibinst.mit.edu/BURNDY/Collections/Babson/OnlineNewton/NewtoniaPDF/PrincipiaMathematica/PrincipiaMathematical687.pdf$

- The text on Modern Algebra by Garret Birkhoff and Saunders Mac Lane, Macmillan, 1941
- Lectures on Classical Differential Geometry by Dirk Struik, Addison-Wesley, 1950
- Principles of Mathematical Analysis as his book was called when it was published by McGraw-Hill in 1953
- Emil Artin and Emma Nöther on Moderne Algebra taken by Bartel van der Waerden, Springer-Verlag, 1940
- Edward Titchmarsh then determined the direction of my efforts: Introduction to the Theory of Fourier Integrals, Eigenfunction Expansions Associated with Second-Order Differential Equations, and The Theory of the Riemann Zeta-Function, Clarendon Press, Oxford, 1937, 1946, and 1951
- Hilbert Spaces of Entire Functions, Prentice-Hall, 1968

[Informal]

- The discovery of the gamma function in 1730
- discovery of the zeta function in 1737
- The functional identity for the zeta function was discovered by Euler only in 1761
- A first step was made in 1811 by Gauss
- The statement of the conjecture in 1859; Riemann hypothesis for the classical zeta function was stated by Bernhard Riemann in 1859

⇒ Monatsberichte der Berliner Akademie, November 1859, 671-680

 $\bullet~$ The interest of Hilbert in the Riemann hypothesis is attested by his 1900 Congress address.

[Colloqium proceedings]

- Colloquium Publication of Raymond Paley and Norbert Wiener, Fourier Transforms in the Complex Domain, American Mathematical Society, 1934
- A source of their work is the Colloquium Publication of Norman Levinson on Gap and Density Theorems, American Mathematical Society, 1940

[Articles - in complete works]

in 1814 by Carl Friederich Gauss, "Methodus nova integralium valores per approximationen inveniendi," Werke, Königliche Gesellschaft der Wissenschaften, Göttingen, 1886, volume 3, pp. 163-196.

⇒ GIDZ

⇒ http://l34.76.163.65/servlet/digbib?template=view.html&id=137431&startpage=171&endpage=204&image=path=http:
//134.76.176.141/cgi-bin/letgifsfly.cgi&image-subpath=/3600&image-subpath=3600&pagenumber=171&imageset-id=3600

- in 1730 by Leonard Euler, "De progressionibus transcendentalibus seu quarum termini generales algebraice dari nequeunt," Opera Omnia I (14), 1-24.
- in 1737, "Variae observationes circa series infinitas," Opera Omnia I (14), 216-244.
- Euler discovered for the classical zeta function in 1761 : "Remarques sur un beau rapport entre les séries de puissances tant directes que réciproques," Opera Omnia I (15), 70-90.

⇒ Gallica? not yet! ⇒ http://vax.wcsu.edu/~sandifer/Beaurapp/Beaurapp.htm

- Erhard Schmidt also made a contribution to the theory of integral equations, "Entwicklung willkürlicher Funktionen nach Systemen vorgeschriebener," Dissertation, Göttingen, 1905
- the author in his thesis, "Local operators on Fourier transforms," Duke Mathematical Journal 25 (1958), 143-153