EuDML | The EUROPEAN DIGITAL MATHEMATICS LIBRARY

ICTPSP

DELIVERABLE

Project Acronym: EuDML

Grant Agreement number: 250503

Project Title: The European Digital Mathematics Library

Deliverable 3.1 - Report on available collections and metadata

Revision: 1.6

Authors:

Thierry Bouche (UJF/CMD)

Hugo Manguinhas (IST)

Main contributors:

Claude Goutorbe, Nicolas Houillon (UJF/CMD)
Rosa de la Viesca, Helena Fernandez (CSIC)
Marie-Louise Chaix, Jean-Paul Jorda (EDPS)
Michael Jost (FIZ); Wojtek Sylwestrzak, Katarzyna Zamlynska (ICM)
Radoslav Pavlov (IMI-BAS)
Jiří Rákosník (IMAS)
Petr Sojka, Miroslav Bartošek (MU)
Ioannis Karydis (IU)
Thomas Fischer (SUBGoe)

Project co-funded by the European Commission within the ICT Policy Support Programme		
Dissemination Level		
P	Public	
C	Confidential, only for members of the consortium and the Commission Services	PP

Revision History

Revision	Date	Author	Organisation	Description
1.0	20/07/2010	Th. B. & H. M.	UJF/CMD & IST	First consolidated version
1.1	22/07/2010	Th. B.	UJF/CMD	Input from IST and MU
1.2	23/07/2010	Th. B.	UJF/CMD	Input from J. Rákosník
1.3	24/07/2010	Th. B.	UJF/CMD	Input from P. Sojka and ML Chaix
1.4	05/08/2010	J. R.	IMAS	New template used
1.5	05/08/2010	J. B.	IST	General revision
1.6	05/08/2010	J. R.	IMAS	Final revision

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Document Index

1. Introduction

The EuDML project aims to design and build a collaborative digital library service that will collate the mathematical content brought by 11 of its partners and make it accessible from a single platform, tightly integrated with relevant infrastructures such as the Zentralblatt MATH. As such, it is the first attempt toward a large-scale implementation of a Digital Mathematics Library (DML), and is expected to pave the way towards a truly inclusive and global DML

Integration of the collections will be an iterative process. It will start by just registering basic metadata for each contributed item (a book, a journal article, a contribution to a collective work such as conference proceedings, a Ph.D. Thesis, etc.), together with a unique identifier. Next steps will use the outcome of a number of work packages to endow each contributed item with a sufficient number of fields ensuring effective exploitation of the whole aggregated content (seamless navigation, interlinking, language-neutral and math-aware retrieval, accessibility, etc.).

1.1. The purpose of work package 3

Work package 3 of the EuDML project (WP3) is tasked with:

1. Reviewing the different metadata schemas in use by partners;
2. Developing a common metadata schema providing enough room to store existing metadata as well as enhancements foreseen by the project;
3. Facilitating the mapping of existing metadata to the common metadata standard;
4. Supporting the conversion and harvest of providers' metadata into the central EuDML registry.

WP3 also works together with other work packages in the project. In particular WP3 works closely with WP4 where the overall system architecture is being designed. It is also in continuous communication with WP6-10, feeding information about available metadata for their work, ingesting new or refined metadata they will generate, structuring the common format in such a way that bottlenecks affecting features or performance of the delivered EuDML system are avoided.

1.2. Overview of the deliverable

This deliverable is the first outcome of this work and is based on a survey of the content that partners contracted to provide to the EuDML project. These collections are outlined in the Annex I (Description of Work, DoW) of the project (pp. 11–14). The purpose of this document is to give a much more precise picture of the available items in the collections, and their associated metadata (both in terms of structure, aka schema, and shape, aka encoding or serialization). A review of the OAI-PMH interoperability among our partners is included as well, so that a comprehensive knowledge base is now available, documenting the state-of-the-art and issues to be tackled towards building the EuDML central metadata repository.

The second section summarizes the findings of this study.

The third defines some basic concepts, a familiarity with these will be useful for the rest of the deliverable.

Next we give a list of the partners and of those collections they contracted to contribute to the EuDML project, which is augmented by our associated partner SUB Göttingen who provides 2 major collections.

Then, for each collection, we survey the size and nature of the corpus, the metadata existing currently, as well as interoperability devices.

A report on OAI-PMH deployment follows.

For the sake of completeness, we analyze the various metadata schemas that were encountered during the study, namely:

- CEDRAM DTD
- DML-E database structure
- EDPS Publishing DTD
- ELibm XML Schema
- ZMATH XML Schema
- BWmeta DTD

For practical reasons these schemas are not included in this document, but they are available to the project and can be provided to the outside at request.

The last two pages of this material contain the Content provider questionnaire.

2. Executive summary

This survey gives us a clearer picture of the content actually brought to the project by its partners.

2.1. Content quantity

Regarding the size of the corpus, good news are that many collections are still growing, both in number of items and type of publications, so that figures can be substantially higher today than when the DoW was written. We refined the notions of digital 'objects' and 'items' to be able to distinguish between existing holdings in libraries to which some metadata is attached (e.g. digitised journal physical issue), and relevant logical units which form the content to be delivered through EuDML (e.g. one multivolume textbook, one journal article).

The EuDML partners bring today 234 000 items, with an expected addition of 21 000 till the end of the project.

2.2. Metadata quality

Regarding the metadata, it was amazing to confirm how much, although general standards are well known and mastered by all partners, each individual project has developed its own dialect for its own purpose, yielding even conflicting uses of very basic schemas such as simple Dublin Core. For ease of analysis, we introduced the notion of 'basic' metadata which provides the minimal bibliographic information allowing building a reasonable digital library service with browsing, searching and matching interfaces. We called 'subbasic' all metadata that do not qualify as 'basic' (e.g. metadata for a journal issue which misses the article-level information). One of the purpose of several work packages in the EuDML project will be to generate and exploit more 'advanced' metadata, which is defined as anything beyond basic in the scope foreseen by the project (multilingualism, MathML formulae, links to reviewing databases, linked bibliographic references, etc.). Of course, this is a rather vague category, as collections with 'advanced' metadata differ widely in the precise 'advancements' they enjoy, and none of them has all of those we would like to endow them with as an outcome of the project. Nevertheless, the summarised figures of 44% basic, and 55% advanced items' metadata give a rough picture of the work to be carried out in WPs 6-10. Only 1% items have metadata categorized as subbasic, but beyond them there are quite a number of interesting resources which should be eventually integrated, like Gallica journals for which only subbasic metadata is available. Another indication of quality is that 44% of our items have an associated searchable full text.

2.2.1. Interoperability

It was found that, among the 13 contributed collections, 10 have an OAI-PMH server already running, but only 4 out of them can be used in their current state in order to bring their content to EuDML central registry (in fact, all but one among these still need parsing in order to reconstruct item-level clean data). Interoperability is thus an area where a certain effort is required. Some details on the work to be done in this respect have been collected in the last section of this report.

2.2.2.Collections summary

Provider/Collection	subbasic	basic	advanced[1]	Searchable full text	OAI-PMH server[2]	EuDML-ready metadata served
CMD/CEDRAM			1 297	1 297	●	No
CMD/NUMDAM			40 478	40 478	●	No
CSIC/DML-E			6 401			No
EDPS journals			2 723		●	Yes
FIZ/ElibM	2 000	7 000	35 000	35 000		No
ICM/DML-PL		13 077	400			No
IMI-BAS-BulDML	270			270	○	No
IMAS/DML-CZ		1 737	23 734	25 471	○	No
IU/HDML		8 241			○	No
SUBGoe Mathematica		69 281			●	Yes
SUBGoe/RusDML			16 748		●	Yes
BNF/Gallica-Math		4 070			○	No
BNP/Port. Mat.			1 347		●	Yes
Total	2 270	103 406	128 128	102 516		
FIZ/ZMATH			2.9 million			

[1] The count of advanced items is somewhat overestimated as it is based in many cases on the standards of the collection that host them rather than on individual item scrutinizing, which was not always possible in the scope of this study. As a result, all items in collections with advanced standards have at least basic metadata, and advanced details such as an MR or ZMATH ID whenever relevant to that item.

[2] A bullet means: yes, with basic metadata served, a circle means: yes, with subbasic metadata served.

3. Preliminaries

3.1. Some basic concepts

The EuDML project aims at integrating mathematical content from a variety of providers across Europe. The selection criterion of this content is analogous to that used by mathematical libraries in universities: it should be a published mathematical work, scientifically validated, in final form, so as to be ready to serve as reliable reference for any scientist from high school pupils to leading-edge researchers. Moreover, it should exist in digital form in order to be contributed to the EuDML service. Each partner contributing some content to the EuDML project has a repository of digital files (typically: PDF) holding published mathematics, and an information system enabling its management and exploitation.

In the scope of this deliverable, a digital object is any such file, to which some metadata is associated so that the file can be handled (curated, retrieved, served over the internet, etc.).

For instance, the digitisation of one book can result in the creation of a large number of digital objects, such as: each page, each chapter, and the book itself. The metadata of a page is typically its file name and location, the ordering in the linear sequence of physical pages forming the book, and possibly the logical page number attached to it, as well as other information such as orientation, or an OCRed version of the content of that page. The metadata attached to a chapter might be the ordered collection of pages constituting this chapter, its title. The principal metadata of the whole book would be its title, author, publisher, year and place of publication, ISBN, the URL of the actual PDF, etc.

Among these digital objects, we identified those which are relevant logical units to be delivered in the context of a scientific library, which we will call items in the sequel.

An item is a self-contained mathematical text which has been scientifically validated and formally published. A monograph is an item, its chapters may be considered items, but this is not considered critical; a journal article is an item rather than a full journal issue. A proceedings volume or any collective book may be an item, as well as each of its individual contributions. A multiple volume work is one item if it is one single work which has been output in smaller units just for handling convenience.

As the main focus of the EuDML project will be to ease retrieval, access and use of mathematical items, we will analyse existing metadata from this perspective.

An item has basic metadata if the metadata identifies it unambiguously not only among the contributor's items, or even EuDML items, but among the whole literature. For a book, it means that it stores at least authors' names, title, publisher, edition, and date of publication (and the fact that it's a book). For a journal article, it stores the journal title, ISSN if available, volume, issue, article title, authors' names, page numbers and any article identifier system relevant for journals that have all articles starting on page 1 (and the fact that it's a journal article). Basic metadata is the requirement for setting up a reasonable browsing of the collections (by authors, title, years, series, volumes), as well as for matching a written reference with an item. Keywords, subject classifications, abstracts are appreciated components of basic metadata, although not mandatory to be qualified as such.

An item has subbasic metadata if it is not sufficient to reconstruct its basic bibliographic reference so as to locate it in a paper library, e.g., or to identify it without ambiguity. A typical example of subbasic metadata is the pair: title/author, possibly with the mention of a source such as journal or a publisher: this sometimes identifies some books, but could not for instance distinguish this book from a short announcement of its content published by its author under the same title.

What we will call advanced metadata is much less constrained, as it essentially means anything beyond basic that is meaningful for our corpus and user base. This ranges from tagged references lists, links to relevant resources or ontologies, multilingual metadata (such as translated or transliterated titles, keywords, and abstracts), MathML formulae, etc.

In this survey, we also investigated the existence of searchable full texts, which we consider as a special kind of advanced metadata. For some purely textual resources, one might wonder whether the full text is still a metadata rather than the data itself. A well-structured full text can make it possible to derive from a unique source both presentation formats and searching-oriented metadata. In the case of mathematical content, there is no such format: one of the endeavors of the EuDML project will be to produce textual, machine-readable versions of the full texts so as to enhance user experience and accessibility to the very mathematical meaning borne by the corpus. A searchable full text is thus to be understood as some textual approximation to the actual content of an item. It may be extracted from a PDF, generated by OCR, manually keyed, converted from TeX, represented by structured XML with MathML formulae.

3.2. Methodology

In order to conduct promptly this survey, a call was issued to all EuDML content providers at the Lisbon kick-off meeting in February 2010. Some pages were set-up in the wiki of the project (wiki.eudml.eu) where each partner could input a description of its metadata formats, and a sample of its holdings. For a more global overview of size, nature, and interoperability of the collections a questionnaire (see Chapter 8. Appendix) was sent through the project's mailing list. It is expected that this questionnaire will become part of the transactions with future associated partners willing to contribute more collections.

A number of reminders followed, and some more informal bilateral communication in order to achieve a precise picture of each collection.

In order to perform an objective study of partner's holdings, and as a contribution to Task 3.3 (designing the EuDML schema), and to the design and implementation tasks in WP4 and WP5, an instance of the REPOX metadata manager was installed and used to harvest to the extent possible all the OAI servers known to the project. The metadata obtained through various channels (sample XML files or database dumps, various formats harvested) was inspected in detail. As it has been envisaged in Task 3.3 to base the EuDML metadata schema on the Journal Archiving and Interchange Tag Set Tag Library of the US National Institutes of Health (NIH), partner's XML files were converted to this format so as to simultaneously assess whether the schema was versatile enough to store faithfully all structures we encountered, and to get a deep understanding of the available tags and their use in EuDML collections.

4. EuDML Content Providers

4.1. EuDML partners

Acronym	Name	Collections	Country	Website
CMD	Cellule MathDoc (UJF/CNRS)	NUMDAM CEDRAM	FR	http://www.mathdoc.fr/
CSIC	Instituto de Estudios Documentales sobre Ciencia y Tecnología — IEDCYT	DML-E	ES	http://www.cindoc.csic.es/
EDPS	EDP Sciences	Math. journals	FR	http://publications.edpsciences.org/
FIZ	Fachinformations-zentrum Karlsruhe, Zentralblatt MATH	ELibM Zentralblatt MATH	DE	http://www.zentralblatt-math.org/
ICM	Interdisciplinary Centre for Mathematical and Computational Modeling, University of Warsaw	DML-PL	PL	http://www.icm.edu.pl/
IMAS	Matematický ústav AV ČR, v.v.i.	DML-CZ	CZ	http://www.math.cas.cz/
IMI-BAS	Institute of Mathematics and Informatics – Bulgarian Academy of Sciences	BulDML	BG	http://www.math.bas.bg/
IU	Ionian University: Department of Informatics	HDML	GR	http://www.ionio.gr/depts/cs/

4.2. Associated partner

Acronym	Name	Collections	Country	Website
SUBGoe	Göttingen Digitization Centre at Niedersächsische Staats- und Universitätsbibliothek	Mathematica RusDML	DE	http://gdz.sub.uni-goettingen.de/

4.3. Content providers through an EuDML partner

Acronym	Name	Collections	Partner	Country	Website
BNF	Bibliothèque nationale de France	Gallica-MATH	CMD	FR	http://www.bnf.fr/
BNP	Biblioteca Nacional de Portugal	Digitized *Portugaliae Mathematica*	IST	PT	http://www.bnp.pt/

5. Provided collections

5.1. CMD: CEDRAM

Collection Description	
Name	Centre de diffusion de revues académiques mathématiques (CEDRAM)
Description	Born digital articles from 5 journals and 5 seminar proceedings
Size	10 periodicals, **1297 articles**
Expected growth	250 items/year
Owner	CNRS/UJF and publishers

Metadata Repository	
Software	EDBM
Internal Schema	CEDRAM DTD
Metadata	Advanced + searchable full texts (pdftotext and LaTeX)
Text encoding	Unicode (UTF-8)
Math encoding	LaTeX and MathML
Service Interface	OAI-PMH
OAI Request URL	http://www.cedram.org/oai
OAI Sets	One per journal
Supported Schema(s)	OAI-DC, mini-DML
Access	Open Access
Notes	Mostly a superset of NUMDAM metadata, CEDRAM's is derived from LaTeX source using Tralics. More than a half items have structured bibliographies, bilingual abstract and keywords, MSC. All math metadata is dually encoded in MathML and in LaTeX.

User access to the content	
Description	Browsing and searching, links
Access	7 series Open access; 1 (resp. 2) with 1 (resp. 5) year moving wall
URL	http://www.cedram.org/
Persistent Identifiers	Yes (internal schema)
	http://ccirm.cedram.org/item?id=CCIRM_2010__1_2_1_0
Format	PDF

Service access to the metadata by EuDML central registry	
Description	Open access to advanced descriptive metadata
Service Interface	OAI-PMH
Metadata Schema	NLM Journal archiving

5.2. CMD: NUMDAM

Collection Description	
Name	Numérisation de documents anciens mathématiques (NUMDAM)
Description	Mostly digitised articles from French journals (1810-2005). But also born digital articles, seminar proceedings, Ph. D. Thesis, and research monographs
Size	30 journals, 29 seminars (4869 volumes), 270 Ph. D. Thesis, 1 series of monographs (165 books). **40478 items**
Expected growth	6 journals full runs: 5000 items, born digital acquisition from publishers: 300 items/year
Owner	CNRS/UJF and many publishers

Metadata Repository	
Software	EDBM
Internal Schema	NUMDAM DTD
Metadata	Advanced + searchable full texts (OCR/pdftotext)
Text encoding	Unicode (UTF-8)
Math encoding	LaTeX in titles, flat Unicode elsewhere
Service Interface	OAI-PMH
OAI Request URL	http://www.numdam.org/oai
OAI Sets	One per series (59)
Supported Schema(s)	OAI-DC XML, mini-DML
Access	Open Access
Notes	NUMDAM has tagged bibliographic references, links to Math Reviews and Zentralblatt MATH for articles and citations, multilingual metadata in some cases, abstract, keywords, MSC in some cases.
User access to the content	
Description	Browsing and searching, links
Access	43 series Open access; 16 with moving wall (generally 5 years)
URL	http://www.numdam.org/
Persistent Identifiers	Yes (internal schema)
	http://www.numdam.org/item?id=AMPA_1810-1811__1__5_0
Format	PDF, DjVu

Service access to the metadata by EuDML central registry	
Description	Open access to advanced descriptive metadata
Service Interface	OAI-PMH
Metadata Schema	NLM Journal archiving

5.3. CSIC: DML-E

Collection Description	
Name	Biblioteca Digital Española de Matemáticas (DML-E)
Description	Digitised and born digital articles from journals published in Spain
Size	22 journals, **6246 articles**
Expected growth	Unknown
Owner	CSIC and the respective publishers

Metadata Repository	
Software	Unknown
Internal Schema	DML-E (comparable to simple DC)
Metadata	Advanced
Text encoding	Unicode (UTF-8)
Math encoding	Unicode/HTML
Service Interface	None
Notes	English translation of title and abstract, UNESCO subject classification, links to MR and ZMATH

User access to the content	
Description	Browsing and searching
Access	15 journals Open Access, 7 with moving wall (1-5 years)
URL	http://dmle.cindoc.csic.es/
Persistent Identifiers	Unknown
Format	PDF

Service access to the metadata by EuDML central registry	
Description	No service interface is currently available. OAI-PMH harvesting will be considered first choice.
Service Interface	To be decided
Metadata Schema	To be decided (probably NLM journal archiving)
Notes	The full database was transferred to CMD, where records will be exported through OAI-PMH.

5.4. EDPS: mathematical journals

Collection Description	
Name	EDP Sciences mathematical journals
Description	Born digital articles published by EDP Sciences since 1997
Size	7 journals, **2723 articles**
Expected growth	350 items/year
Owner	EDPS

Metadata Repository	
Software	In-house Software
Internal Schema	EDPS Publishing DTD
Metadata	Advanced
Text encoding	Unicode (UTF-8)
Math encoding	LaTeX
Service Interface	OAI-PMH
OAI Request URL	http://oai.edpsciences.org/
OAI Sets	One per journal.
Supported Schema(s)	OAI-DC, EDPS Publishing DTD
Access	Open access
Notes	Structured and tagged bibliographies, author affiliations, English translations

User access to the content	
Description	Browsing and searching
Access	Moving wall (max. 5 years, 50% items open access)
URL	http://publications.edpsciences.org/
Persistent Identifiers	Yes (DOI)
	http://dx.doi.org/10.1051/m2an:2008032
Format	PDF

Service access to the metadata by EuDML central registry	
Description	Open access to all metadata, not full texts
Service Interface	The main interface will be OAI-PMH. Data can also be sent using HTTP or FTP if needed.
Metadata Schema	EDPS Publishing DTD

5.5. FIZ: ELibM

Collection Description	
Name	The Electronic Library of Mathematics (ELibM)
Description	Born digital articles (mostly "Open Access – Gold") from a variety of publishers
Size	91 current journals (new articles arriving) + 13 archival (no new articles arriving), about 44000 articles
Expected growth	5% per year
Owner	FIZ and the respective publishers
Notes	ELibM is itself a library that mirrors production from third party publishers. Formats and metadata are thus very heterogeneous.

Metadata Repository	
Software	Various
Internal Schema	ELibM
Metadata	About 9000 (sub)basic, 35000 advanced + searchable full texts
Math encoding	TeX
Text encoding	Unicode (UTF-8)
Service Interface	None
Notes	Experimental ESciDoc repository under construction. Metadata availability is often decoupled from publishing of articles, since most publishers don't supply standardized interoperable metadata.

User access to the content	
Description	Browsing and searching
Access	Open Access and moving wall
URL	http://www.emis.de/journals/
Persistent Identifiers	None
Format	PDF, DVI, PS, depending on publisher

Service access to the metadata by EuDML central registry	
Description	OAI-PMH harvesting will be considered first choice. Alternatively, metadata in XML files
Service Interface	None yet. OAI-PMH interface will have to be installed.
Metadata Schema	Yet to be decided. Probably native ELibM metadata converted to suitable schema.

5.6. FIZ: Zentralblatt MATH

Collection Description	
Name	Zentralblatt MATH (ZMATH)
Description	Reviewing database covering the world mathematical output from 1868 onward
Size	About 3 million reviews
Expected growth	120 000 reviews/year
Owner	FIZ, Heidelberger Akademie der Wissenschaften, EMS, Springer
Notes	The content in ZMATH is metadata only, with (external) links to full text where available. EuDML collections represent less than 9% of ZMATH content.

Metadata Repository	
Software	EDBM and in-house solutions
Internal Schema	ZMATH Schema (in-house)
Metadata	advanced
Text encoding	Unicode (UTF-8)

Math encoding	(plain) TeX, MathML (derived from TeX source)
Service Interface	None
Notes	Access to contents is granted to EuDML project partners (file transfer).

User access to the content	
Description	Advanced retrieval interface supporting multiple result list display options. DOI or other PURL linking to (external) full text where available, simple URL linking otherwise
Access	Full access to ZMATH database is restricted to subscribers; free evaluation access available by registration; free limited trial access available to all.
URL	http://www.zentralblatt-math.org/zbmath/
Persistent Identifiers	Yes (Zbl IDs)
	http://www.emis.de/zmath-item?1138.14001
Format	XHTML, MathML, XML, ASCII, BibTeX, PDF
Notes	Access to indexed full texts is under the control of their respective publishers. Fields above describe access to ZMATH content (reviews and other metadata).

Service access to the metadata by EuDML central registry	
Description	Subset of ZMATH metadata (related to partner collections) will be available to EuDML central registry. Details to be decided.
Service Interface	To be decided
Metadata Schema	To be decided

5.7. ICM: DML-PL

Collection Description	
Name	DML-PL (formerly known as BWM—Biblioteka Wirtualna Matematyki)
Description	Retrodigitised and born digital articles and book chapters from 14 series published in Poland
Size	10 journals, 4 series of books, **13477 items**
Expected growth	1600 articles added next years
Owner	ICM
Notes	BWM is currently ongoing upgrade to Yadda-based DML-PL

Metadata Repository	
Software	YADDA
Internal Schema	BWmeta
Metadata	Basic (13077 items), advanced (400 items)
Text encoding	Unicode (UTF-8)
Math encoding	Unicode/TeX/MathML

Service Interface	None
Notes	BWM metadata was basic, the DML-PL metadata, which currently mostly concerns born digital items, is advanced, with translations of non-English metadata, and tagged bibliographies. The structure of BWmeta DTD, where for instance citation's titles are stored as attributes may limit the possibility to deal with, e.g. MathML.

User access to the content	
Description	Browsing and searching (Yadda)
Access	Open Access
URL	http://matwbn.icm.edu.pl/ (BWM) http://yadda.icm.edu.pl/mathbwn/ (prototype DML-PL)
Persistent Identifiers	Yes
Format	PDF

Service access to the metadata by EuDML central registry	
Description	No service interface is currently available. OAI-PMH harvesting will be considered first choice.
Service Interface	To be decided
Metadata Schema	To be decided

5.8. IMI-BAS: BulDML

Collection Description	
Name	Bulgarian DML (BulDML)
Description	Born digital articles from *Serdica Mathematical Journal* (1995–2002) and *Serdica Journal of Computing* (2007–2009)
Size	2 journals, **270 articles**
Expected growth	4 journals full runs, born digital as well as retrodigitised (1605 items)
Owner	IMI-BAS

Metadata Repository	
Software	Dspace
Internal Schema	DSpace DBMS Schema (with prescribed metadata schemas for BulDML)
Metadata	Subbasic + searchable full texts
Text encoding	Unicode (UTF-8)
Math encoding	Unicode
Service Interface	OAI-PMH
OAI Request URL	http://sci-gems.math.bas.bg:8080/oai/request
OAI Sets	*Serdica Mathematical Journal:* hdl_10525_396 *Serdica Journal of Computing:* hdl_10525_3

Supported Schema(s)	OAI-DC XML, METS with embedded MODS
Access	Open Access

User access to the content	
Description	Browsing and searching (Dspace)
Access	Open Access
URL	http://sci-gems.math.bas.bg:8080/jspui/handle/10525/2
Persistent Identifiers	Yes (handle)
	http://hdl.handle.net/10525/613
Format	PDF

Service access to the metadata by EuDML central registry	
Description	Open access to descriptive metadata
Service Interface	OAI-PMH
Metadata Schema	METS with embedded MODS
Notes	The METS format qualifies all repeated elements. It is thus the most precise harvesting format. However it lacks volume, issue and page numbers (volume and issue can be reconstructed from the set spec, pages seems to be absent from the BulDML metadata).

5.9. IMAS & MU: DML-CZ

Collection Description	
Name	Czech Digital Mathematics Library (DML-CZ)
Description	Retrodigitised and born digital articles, books, and proceedings published in Czech lands
Size	11 journals, 6 proceeding series, 35 books including complete works of Bernard Bolzano; **25471 items** in total (20000 of which retrodigitised)
Expected growth	1 journal full run, 1 monograph series (40 books), born-digital acquisitions from publishers (total 4000-5000 new items in the next 2 years); update frequency: monthly
Owner	IMAS

Metadata Repository	
Software	DSpace
Internal Schema	Dspace DBMS Schema (with prescribed metadata schemas for DML-CZ)
Metadata	Advanced + searchable full texts
Text encoding	Unicode (UTF-8)
Math encoding	TeX
Service Interface	OAI-PMH
OAI Request URL	http://oai.dml.cz/request

OAI Sets	One per journal issue (2316 sets) as implicit in DSpace, could be changed
Supported Schema(s)	OAI-DC, QDC, MODS, METS, DIDL
Access	Open Access
Notes	Advanced features: reference lists, translated titles (into English), MSC, links to MR/ZMATH, similar articles. Some advanced metadata fields (subject; bibliographic references) are exposed by OAI-PMH in OAI_DC and DIDL formats, but basic ones such as volume number or page numbers are stored in an unqualified description element.

User access to the content	
Description	Browsing and searching (DSpace/Lucene), links to MR/ZM and similar articles, dynamic visualisation/browsing in development (Visual Browser), mathematical formula search in development.
Access	Open access (more than 96% items) and moving wall (for 6 of total 11 journals, max. 2 years)
URL	http://dml.cz/
Persistent Identifiers	Yes (handle)
	http://hdl.handle.net/10338.dmlcz/123200
Format	PDF 1.4
Notes	Indexed by Google by special agreement, 60%+hits from Google. PDFs are digitally signed and thus distinguishable from other copies. PDF recompression (40% of original size) in preparation.

Service access to the metadata by EuDML central registry	
Description	Open access to advanced metadata
Service Interface	OAI-PMH; data can also be sent using HTTP or FTP if needed
Metadata Schema	To be defined. Probably METS with embedded metadata schemas for descriptive metadata (OAI-DC or QDC considering the internal schema used for descriptive metadata).

5.10. IU: HDML

Collection Description	
Name	Hellenic Digital Mathematics Library (HDML)
Description	Mostly retrodigitised articles from many mathematical journals published in Greece
Size	8 journals, 29 conference proceedings, 20 books, **8241 items**
Expected growth	Goal is 7 more journals, 7000 items more
Owner	IU, Hellenic Mathematical Society
Notes	http://www.hdml.gr inaugurated operation on 4th of July 2010

Metadata Repository	
Software	Greenstone (http://www.greenstone.org/)
Internal Schema	GAF DTD
Metadata	basic
Text encoding	Unicode (UTF-8)
Math encoding	None
Service Interface	OAI-PMH
OAI Request URL	http://www.hdml.gr/greenstone/cgi-bin/oaiserver.cgi
OAI Sets	hdml
Supported Schema(s)	OAI-DC
Access	Open Access
Notes	Highly multilingual, a lot of educational material

User access to the content	
Description	Greenstone installation
Access	Open Access
URL	http://www.hdml.gr/greenstone/cgi-bin/library.cgi
Persistent Identifiers	Yes
Format	PDF
Service access to the metadata by EuDML central registry	
Description	Open access to all metadata
Service Interface	OAI-PMH
Metadata Schema	GAF DTD
Notes	Not fully active at deliverable write-up

5.11. SUBGoe: Mathematica

Collection Description	
Name	Mathematica collection at Göttinger Digitalisierungszentrum
Description	Retrodigitised journals and books
Size	42 journals (2547 volumes), 1531 monographs, 294 multi-volume books (742 volumes), **69281 items**
Expected growth	Still growing
Owner	SUBGoe
Notes	5 journals are duplicated at DML-CZ with advanced metadata there, 1 is a digitized copy of Zentralblatt

Metadata Repository	
Software	unknown
Internal Schema	METS profile with embedded descriptive metadata encoded in MODS and rights related metadata encoded in DFG-viewer schema

Metadata	Basic
Text encoding	Unicode (UTF-8)
Math encoding	None
Service Interface	OAI-PMH
OAI Request URL	http://gdz.sub.uni-goettingen.de/oai2/
OAI Sets	"mathematica" (includes RusDML)
Supported Schema(s)	OAI-DC, METS
Access	Open Access
Notes	Metadata is at the volume level. It is possible to derive basic article-level (or chapter-level) metadata from METS format. For books, beyond basic metadata, a table of contents with chapter details can be built. Documentation of GDZ use of METS format here: http://gdz.sub.uni-goettingen.de/index.php?id=46&L=1 It was not possible to harvest large quantities of records from GDZ' OAI server. An alternative method based on first identifying the PPN of each item, and then downloading it individually was suggested by SUBGoe. It was used successfully to download all journal volumes in this collection.

User access to the content	
Description	Searching
Access	Open Access
URL	http://gdz.sub.uni-goettingen.de/ An alternative access with different searching and browsing interface to a subset of the collection is available from Digizeitschriften http://www.digizeitschriften.de/
Persistent Identifiers	Yes (PPN)
	http://gdz.sub.uni-goettingen.de/index.php?id=resolveppn&PPN=GDZPPN00236736X
Format	PDF, DFG-viewer

Service access to the metadata by EuDML central registry	
Description	GDZ has already a lot of interoperability devices. As the METS format, which is delivered though OAI as well as individually (e.g. http://gdz.sub.uni-goettingen.de/mets_export.php?PPN=PPN509860087_0015), has a complete and faithful representation of GDZ internal holdings, it will not be necessary that SUBGoe sets up a dedicated service for EuDML to exploit its metadata.
Service Interface	For small amount of metadata, such as incremental daily update, OAI-PMH with METS format seems appropriate. For the large initial download, a different method using the various existing interfaces has to be set up.
Metadata Schema	METS Profile

5.12. SUBGoe: RusDML

Collection Description	
Name	RusDML
Description	Retrodigitised and born digital journal articles from Russia through a cooperation with German partners (FIZ, SUBGoe, Hannover…)
Size	11 journals (603 volumes), **16748 items**
Expected growth	Project finished in principle
Owner	SUBGoe (with Zentralblatt, and the Russian academy of science)

Metadata Repository	
Software	Unknown
Internal Schema	Same METS profile used in Mathematica collection from SUBGoe.
Metadata	Advanced
Text encoding	Unicode (UTF-8), transliterated Cyrillic
Math encoding	None
Service Interface	OAI-PMH
OAI Request URL	http://gdz.sub.uni-goettingen.de/oai2/
OAI Sets	“rusdml”
Supported Schema(s)	OAI-DC, METS
Access	Open Access
Notes	Metadata is at the volume level. It is possible to derive article-level metadata from METS format. Some metadata comes from ZM, ZM IDs are provided, as well as translations and transliteration of Cyrillic text and names.

User access to the content	
Description	Searching
Access	Open Access
URL	http://gdz.sub.uni-goettingen.de/
Persistent Identifiers	Yes (PPN)
	http://gdz.sub.uni-goettingen.de/index.php?id=resolveppn&PPN=GDZPPN00236736X
Format	PDF, DFG-viewer

Service access to the metadata by EuDML central registry	
Description	GDZ standard interfaces
Service Interface	Same as in mathematica collection from SUBGoe
Metadata Schema	METS Profile

5.13. BNF: Gallica-MATH

Collection Description	
Name	Gallica-MATH
Description	Item-level catalogue of retrodigitised journal articles and collected works
Size	1 journal (98 volumes), 98 books, **4070 items**
Expected growth	Gallica holds more digital objects in mathematics with basic metadata (817 books) or subbasic (329 periodical volumes) which should be considered.
Owner	BNF

Metadata Repository	
Software	XML files
Internal Schema	Gallica-Math
Metadata	Basic
Text encoding	Unicode (UTF-8)
Math encoding	TeX
Service Interface	OAI-PMH
OAI Request URL	http://oai.bnf.fr/oai2/OAIHandler
OAI Sets	GallicaMath
Supported Schema(s)	OAI-DC, TEL Application Profile. DIDL
Access	Open Access

User access to the content	
Description	Browsing and Searching
Access	Open Access
URL	Resources on http://gallica.bnf.fr/ mediated via http://portail.mathdoc.fr/GALLICA/
Persistent Identifiers	Yes (BNF: ark; CMD: internal schema)
	http://gallica.bnf.fr/ark:/12148/bpt6k26502w/f30n63 http://math-doc.ujf-grenoble.fr/cgi-bin/oeitem?id=OE_KLEIN__3_13_0
Format	PDF

Service access to the metadata by EuDML central registry	
Description	Open access to basic metadata
Service Interface	OAI-PMH
Metadata Schema	Mini-DML and/or NLM archiving

5.14. BNP: Portugaliae Mathematica

Collection Description	
Name	Portugaliae Mathematica
Description	Retrodigitised journal articles
Size	1 journal (200 volumes), **1347articles**
Expected growth	Finished in principle (some newer volumes are in ELibM, the latest ones being published by EMS-ph)
Owner	BNP, SPM

Metadata Repository	
Software	Repox
Internal Schema	BNP Schema
Metadata	Advanced
Access	Open Access
Text encoding	Unicode (UTF-8)
Service Interface	OAI-PMH
OAI Request URL	http://oai.bn.pt/servlet/OAIHandler
OAI Sets	PortugalMatematica
Supported Schema(s)	OAI-DC, XArq, TEL Application Profile, MODS, Agrisap, EAD.
Math encoding	TeX
Notes	Links to MR/ZM. Current OAI-DC format represents bibliographic information as a string.
User access to the content	
Description	Searching and browsing
Access	Open Access
URL	http://purl.pt/index/pmath/PT/index.html
Persistent Identifiers	Yes (PURL)
	http://purl.pt/2603
Format	PDF

Service access to the metadata by EuDML central registry	
Description	Open access to advanced metadata
Service Interface	OAI-PMH
Metadata Schema	BNP Schema.

5.15. Prospective content partners

Beyond the EuDML grant agreement beneficiaries and associated partners listed above, there are a number of projects active in Europe, at various development stages, which we will try to associate to our effort. We give here a short informal account on their holdings and the interoperability they offer.

5.15.1. SwissDML

SwissDML is a service of a consortium of Swiss university libraries, part of Swiss Electronic Academic Library Service (SEALS). It currently holds 5186 items from 4 journal series published in Switzerland. Among them, 2 journals had been digitised by SUBGoe up to 1996. Metadata is at best basic. An OAI-PMH server is alive at http://retro.seals.ch/oai/dataprovider but there is no way to extract the page numbers of articles, e.g.

5.15.2. Mathematical Institute of the Serbian Academy of Sciences and Arts

The Mathematical Institute of the Serbian Academy of Sciences and Arts (Belgrade) is interested to become an associated partner of the EuDML project. It currently manages two websites giving access to about 4500 items, and expected to reach more than 5000 during the 2010 summer when a new journal will be included.

The related addresses are http://elib.mi.sanu.ac.rs/ and http://elibrary.matf.bg.ac.rs/ (Virtual library of the Faculty of Mathematics, Belgrade). The last one is using DSpace.

5.15.3. BDIM

BDIM is the Italian project of math digitisation (Biblioteca Digitale Italiana di Matematica). The project has recently started, headed by SIMAI (Societa Italiana di Matematica Applicata e Industriale) and UMI (Unione Matematica Italiana) with initial support from the Biblioteca Digitale Italiana and the Italian Ministry of Beni and Attivita Culturali. At the moment BDIM contains 1814 articles from one journal: *Bollettino dell'Unione Matematica Italiana*, 1946-1967. Metadata standards were derived from NUMDAM's so that integration in EuDML should be straightforward. BDIM has already set up an OAI-PMH server using CMD's mini-DML schema.

6. Service Interfaces

The service interface of choice for contributing metadata to the EuDML central repository is the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Depending on providers' ability to set up an OAI server delivering a suitable exchange format with rich enough metadata, we will test different scenarios such as

- Harvesting through OAI-PMH rich metadata in a format near to the internal format of the provider or easy to deliver within their current set-up.
- Harvesting existing metadata (XML files or database export) through FTP or other file transfer protocol at a partner's site serving as a proxy to EuDML, where files will be suitably transformed and served through OAI-PMH.
- At the end of the project, when the EuDML schema being developed as task 3.3 of this work package will be published, all content providers should have set up a harvesting interface through OAI-PMH serving rich EuDML metadata.

In the end, the central metadata registry will only have to harvest a limited number of servers, and to know the transforming rules to apply to each of the contributed metadata set.

The following partners have already one or more OAI-PMH servers running on top of (some of) their EuDML collections: CMD, EDPS, ICM, IMAS, IMI-BAS, IU, SUBGoe, BNF, BNP. Unfortunately, many of these servers only deliver subbasic metadata (the strict minimum expected from OAI-DC schema, which is: identifier, author, title, date of record, URL of the resource). In fact, our study shows that only in the case of IMI-BAS (METS), SUBGoe (METS) and BNP (OAI-DC), the metadata delivered is faithful to the internal holdings for basic item-level metadata. In order to contribute their collections, all the other partners will have to set up an ingesting path according to one of the above scenarios.

Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH)	
Description	The OAI-PMH provides an application-independent interoperability framework based on metadata harvesting. It specifies a protocol where client applications can collect metadata records from data providers.
Documentation	http://www.openarchives.org/OAI/openarchivesprotocol.html
Latest Version	Version 2.0 (2002-06-14)
Responsible Organization	Open Archives Initiative (OAI)
Communication	HTTP ("GET" requests and XML responses)

7. Information Schemas

This section describes the interchange information schemas that are currently being used by the listed content providers. The information schemas may be divided in Metadata and Content schemas, whether they are used to encode metadata information or content (digital objects).

7.1. Metadata schemas specific to a Data Provider

7.1.1. CEDRAM DTD

CEDRAM DTD	
Description	The CEDRAM schema describes general information about the journal, production process, and volume specific metadata are in a header (element *notice*); A repeatable *article* element provides metadata for each article (IDs, author [full name, first name/last name/vonpart, etc.], page [first, last], title [xml:lang attr, variant with TeX and MathML maths], language (this determines the original title), biblio [bib_entry (with doctype) [bib author, bib title, bib year, etc. (derived from bibtex; there is also a flat version for unstructured bibs)]]).
Metadata Coverage	Descriptive and Structural metadata
Documentation	Available to the project
Responsible Organization	Cellule MathDoc (UJF/CNRS)
Data Format	XML based data format with a schema specified using DTD

7.1.2. NUMDAM DTD

NUMDAM DTD	
Description	The NUMDAM schema describes general information about the journal, digitisation process, and volume specific metadata are in a header (element *notice*); A repeatable *article* element provides metadata for each article (IDs [NUMDAM, MR, ZMATH, JFM], author [first name/last name], page [first, last], title [xml:lang attr], language (this determines the original title), biblio [bibitem [bib author, bib title, bib year, and more]]). Sometimes, non-math articles are also registered in this XML: tables of content, plates, etc.
Metadata Coverage	Descriptive and Structural metadata.
Documentation	Overall architecture shared with CEDRAM DTD
Responsible Organization	Cellule MathDoc (UJF/CNRS)
Schema Type	DTD
Schema Location	http://www-mathdoc.ujf-grenoble.fr/NUMDAM/dtds/volphys.dtd

7.1.3.DML-E DBMS Schema

DML-E DBMS Schema	
Description	Internal schema
Metadata Coverage	Descriptive Metadata
Documentation	Available to the project
Responsible Organization	CSIC
Data Format	DBMS

7.1.4.EDPS Publishing DTD

EDPS Publishing DTD	
Description	Customized versions of the NLM DTD designed mainly to: • Allow the tagging of conference proceedings not published in a journal • Be able to mix presentation and semantic element in the bibliography • Add pub-id-type="bibcode" in article-id tag.
Metadata Coverage	Structural and Descriptive Metadata
Documentation	Available to the project
Responsible Organization	EDPS
Data Format	XML based data format with a schema specified using DTD or XML Schema

7.1.5.ELibM Schema

ELibM Schema	
Description	Schema used for the ELibM digital library
Metadata Coverage	Descriptive metadata
Documentation	Available to the project
Responsible Organization	FIZ
Data Format	XML based data format with no schema specified

7.1.6.ZMATH Schema

ZMATH Schema	
Description	Schema used inside the Zentralblatt MATH Database.
Metadata Coverage	Descriptive metadata
Documentation	Available to the project
Responsible Organization	FIZ
Data Format	XML based data format with no schema specified

7.1.7.BWmeta

BWmeta	
Description	BWMETA is a general-purpose schema capable of representing metadata of entities such as: audio recordings, laws, molecular sequences.
Metadata Coverage	Descriptive and Structural metadata
Documentation	Available to the project
Responsible Organization	ICM
Data Format	XML based data format with no schema specified

7.1.8.BNP Schema

BNP Schema	
Description	Internal schema used for encoding the Portugaliae Mathematica collection
Metadata Coverage	Descriptive (encoded using Simplified DC) and Structural metadata (METS structures)
Documentation	No documentation is available.
Responsible Organization	BNP
Data Format	XML based data format with no schema specified

7.2. Metadata schemas shared among Data Providers

7.2.1.Simplified and Qualified DC

Dublin Core Metadata Element Set	
Description	The Dublin Core Metadata Element Set (DCMES) specifies a list of metadata elements through which resources can be described. The DCMES comes in two flavours, the Simplified and the Qualified Dublin Core (DC). In simplified DC, the range of the metadata elements is not prescribed (normally text expressions are used to encode the values), while in qualified DC, the range is prescribed. Besides the fifteen elements defined in the Simplified DC, Qualified DC includes three additional elements (Audience, Provenance and RightsHolder).
Metadata Coverage	Descriptive Metadata
Documentation	Simplified DC: http://dublincore.org/documents/dces/ Qualified DC: http://dublincore.org/documents/dcmi-terms/
Latest Version	2008-01-14
Responsible Organization	Dublin Core Metadata Initiative
Data Format	RDF encoding formats
Namespace	The namespaces also point to the schema location. Simplified DC: http://purl.org/dc/elements/1.1/ Qualified DC: http://purl.org/dc/terms/

7.2.2.OAI-DC XML

Dublin Core Application Profile for OAI	
Description	OAI-DC XML is an XML format for the serialisation of Simple Dublin Core metadata descriptions. The format is defined as a "metadata format" for use within the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). OAI-PMH requires that data providers support the oai_dc metadata format.
Metadata Coverage	Descriptive Metadata, considered basic for EuDML
Documentation	http://standards-catalogue.ukoln.ac.uk/index/OAI_DC
Latest Version	2002-03-18
Responsible Organization	UKOLN
Data Format	XML based data format specified in XMLSchema (also available as a RDF based format, see OAI-DC RDF)
Namespace	http://www.openarchives.org/OAI/2.0/oai_dc/
Schema Location	http://www.openarchives.org/OAI/2.0/oai_dc.xsd

7.2.3.OAI-DC RDF

Dublin Core Application Profile for OAI	
Description	OAI-DC RDF is an XML format for the serialisation of Simple Dublin Core metadata descriptions.
Metadata Coverage	Descriptive Metadata, considered basic for EuDML
Documentation	http://standards-catalogue.ukoln.ac.uk/index/OAI_DC
Responsible Organization	UKOLN
Data Format	RDF based data model. Can be encoded using the available RDF encoding formats.
Namespace	http://www.openarchives.org/OAI/2.0/rdf/
Schema Location	http://www.openarchives.org/OAI/2.0/rdf.xsd

7.2.4.METS

Metadata Encoding & Transmission Standard	
Description	Standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library. METS is designed for the purpose of: • Creating XML document instances that express the hierarchical structure of digital library objects. • Recording the names and locations of the files that comprise those objects. • Recording associated metadata.
Metadata Coverage	Structure Metadata (with support for embedded Metadata Schema)
Documentation	http://www.loc.gov/standards/mets/ http://www.loc.gov/standards/mets/mets-schemadocs.html

Responsible Organization	The standard is maintained by the Network Development and MARC Standards Office of the Library of Congress, and is being developed as an initiative of the Digital Library Federation.
Latest Version	Current version is 1.9
Data Format	XML based data format specified in XMLSchema
Namespace	http://www.loc.gov/METS/
Schema Location	http://www.loc.gov/standards/mets/mets.xsd

7.2.5.MODS

Metadata Object Description Schema	
Description	A schema for a bibliographic element set that may be used for a variety of purposes, and particularly for library applications. As an XML schema it was originally developed to be able to carry selected data from existing MARC 21 records. It includes a subset of MARC fields and uses language-based tags rather than numeric ones, in some cases regrouping elements from the MARC 21 bibliographic format.
Metadata Coverage	Descriptive metadata
Documentation	http://www.loc.gov/standards/mods/ http://www.loc.gov/standards/mods/mods-overview.html
Responsible Organization	The standard is maintained by the Network Development and MARC Standards Office of the Library of Congress with input from users.
Latest Version	Current version is 3.4
Data Format	XML based data format specified in XMLSchema
Namespace	http://www.loc.gov/mods/v3
Schema Location	http://www.loc.gov/standards/mods/mods.xsd

7.2.6.MathML Schema

Mathematical Markup Language Schema	
Description	MathML is intended to facilitate the use and re-use of mathematical and scientific content on the Web, and for other applications such as computer algebra systems, print typesetting, and voice synthesis. MathML can be used to encode both the presentation of mathematical notation for high-quality visual display, and mathematical content, for applications where the semantics plays more of a key role such as scientific software or voice synthesis.
Metadata Coverage	Encoding of Mathematical related metadata
Documentation	http://www.w3.org/TR/MathML2/
Responsible Organization	W3C Math Working Group as part of W3C Math Activity
Latest Version	Version 2.0 is a W3C Recommendation, released on 21 Feb 2001
Data Format	XML based data format specified in XML DTD
Namespace	http://www.w3.org/1998/Math/MathML
Schema Location	http://www.w3.org/Math/DTD/mathml2/mathml2.dtd

7.2.7.EAD XMLSchema

Encoded Archival Description XMLSchema	
Description	EAD was developed to support the use of holdings related to inventories, registers, indexes, and other documents created by archives, libraries, museums, and manuscript.
Metadata Coverage	Structural and Descriptive metadata
Documentation	http://www.loc.gov/ead/
Responsible Organization	Society of American Archivists and the Library of Congress
Latest Version	200804 Release
Data Format	XML based data format specified in XMLSchema. The schema is also available as RelaxNG.
Namespace	urn:isbn:1-931666-22-9
Schema Location	http://www.loc.gov/ead/ead.xsd

7.2.8.DIDL XMLSchema

Digital Item Declaration Language XMLSchema	
Description	As part of MPEG-21 (ISO/IEC 21000-2:2003), DIDL specifies a uniform and flexible abstraction and interoperable schema for declaring the structure and makeup of Digital Items. Digital Items are structured digital objects, including a standard representation, identification and metadata. They are the basic unit of transaction in the MPEG-21 framework. More concretely, a Digital Item is a combination of resources (such as videos, audio tracks, images, etc), metadata (such as descriptors, identifiers, etc), and structure (describing the relationships between resources).
Metadata Coverage	Structure and Descriptive metadata
Documentation	http://xml.coverpages.org/mpeg21-didl.html http://xml.coverpages.org/MPEG21-WG-11-N3971-200103.pdf
Responsible Organization	Developed by the Moving Picture Experts Group (MPEG), an ISO working group responsible for developing standards for digital audio and video.
Latest Version	2nd version, the first was published in March 2003 and the 2nd in 2005.
Data Format	XML based data format specified in XMLSchema
Namespace	urn:mpeg:mpeg21:2002:02-DIDL-NS
Schema Location	http://download.webct.com/public/ims/2.0/MPEG21.xsd

7.2.9. Mini-DML

Mini-DML	
Description	The mini-DML schema was developed under the Mini-DML project[3] to support the encoding of basic bibliographic data for any kind of mathematical digital article and to support simple search and metadata retrieval.
Metadata Coverage	Descriptive metadata, considered basic for EuDML
Documentation	Not available
Responsible Organization	Cellule MathDoc
Data Format	XML based data format specified in XMLSchema
Namespace	http://www.numdam.org/minidml/elements
Schema Location	http://www.numdam.org/OAI/minidml.xsd

7.2.10. NLM journal archiving DTD

NLM journal archiving and Interchange tag set	
Description	The intent of the NLM Journal Archiving and Interchange Tag Suite is to provide a common format in which publishers and archives can exchange journal content. The Suite provides a set of XML schema modules that define elements and attributes for describing the textual and graphical content of journal articles as well as some non-article material such as letters, editorials, and book and product reviews.
Metadata Coverage	Structural Schema supporting embedded metadata encoded in specific metadata schemas.
Documentation	http://dtd.nlm.nih.gov/archiving/tag-library/
Responsible Organization	National Center for Biotechnology Information (NCBI) National Library of Medicine (NLM) NIH, USA. There is ongoing work at NISO in order to make it a NISO standard, see http://www.niso.org/workrooms/journalmarkup
Latest Version	Current version is 3.0
Data Format	XML based data format specified in XMLSchema
Namespace	http://dtd.nlm.nih.gov/3.0/xsd/archivearticle
Schema Location	http://dtd.nlm.nih.gov/archiving/3.0/xsd/archivearticle3.xsd

7.2.11. DFG Viewer Schema

DFG Viewer Schema	
Description	Defines a set of elements to be used inside METS metadata to encode rights related metadata.
Metadata Coverage	Rights related metadata

[3] http://minidml.mathdoc.fr/

Documentation	http://dfg-viewer.de/fileadmin/groups/dfgviewer/METS_Anwendungsprofil_2.0.pdf
Responsible Organization	The DFG Viewer was essentially developed by SLUB Dresden the suggestion of Deutschen Forschungsgemeinschaft (DFG, German Research Foundation) and on behalf of the four libraries currently involved in the campaign for the digitisation of printed works recorded in national registers (VD16/VD17) with subsidised mass digitisation projects.
Data Format	METS profile.
Namespace	http://dfg-viewer.de/
Schema Location	http://www.loc.gov/standards/mods/mods.xsd

7.2.12. DSpace DBMS Schema

DSpace DBMS Schema	
Description	The DSpace DBMS supports three sorts of metadata about archived content: • The qualified DC is natively supported for the representation of descriptive metadata. Other schemas can also be use for the representation of descriptive metadata which are stored as serialized bitstreams. • Administered metadata is also supported, including preservation metadata, provenance and authorization policy data. Provenance metadata is stored within DC records. • Structural metadata about an item includes information about how to present an item or bitstreams within an item to an end-user, and the relationships between constituent parts of the item. Additional structural metadata can be stored in serialized bitstreams,
Metadata Coverage	Descriptive, Administrative and Structural metadata.
Documentation	http://www.dspace.org/1_6_2Documentation/
Latest Version	Version 1.6.2
Responsible Organization	DuraSpace (http://duraspace.org/)
Data Format	DBMS

7.2.13. GAF DTD

Greenstone Archive Format DTD	
Description	Format used as import format for Greenstone. All documents that are to be ingested into Greenstone must be converted into this format using document processing plugins. This format structures documents into sections, and can hold metadata at the document or section level. Metadata is descriptive information (e.g. author, title, date, keywords) that is associated with a document. This metadata is encoded using specific XML elements and can support different metadata formats like DC.
Metadata Coverage	Descriptive and Structural metadata.
Documentation	http://wiki.greenstone.org/wiki/index.php/Greenstone_Archive_Format

Responsible Organization	Greenstone is produced by the New Zealand Digital Library Project at the University of Waikato, and developed and distributed in cooperation with UNESCO and the Human Info NGO
Data Format	XML based data format with a schema specified using DTD
Namespace	No namespace is declared.
Schema Location	http://www.greenstone.org/dtd/Archive/1.0/Archive.dtd

7.3. Other relevant schemas

7.3.1. MADS

Metadata Authority Description Schema	
Description	MADS is a MARC21-compatible XML format for the type of data carried in records in the MARC Authorities format. As an XML schema for an authority element set that may be used to provide metadata about agents (people, organizations), events, and terms (topic, geographic, genre, etc.). MADS serves as a companion to MODS to provide metadata about the authoritative entities used in MODS descriptions.
Metadata Coverage	Authority metadata.
Documentation	http://www.loc.gov/standards/mads/ http://www.loc.gov/standards/mods/mods-overview.html
Responsible Organization	The standard is maintained by the MODS/MADS Editorial Committee with the Network Development and MARC Standards Office of the Library of Congress and input from users.
Latest Version	Current version is 1.0
Data Format	XML based data format specified in XMLSchema
Namespace	http://www.loc.gov/mads
Schema Location	http://www.loc.gov/standards/mads/mads.xsd

8. Appendix

Following we present the questionnaire sent to the Content Providers.

EuDML | The EUROPEAN DIGITAL MATHEMATICS LIBRARY

Content provider questionnaire

Provider name

Organization's contact person (name, address, email)

Organization's technical contact person (name, address, email)

Collections

Name of each contributed collection, URL where it is currently hosted, and any relevant detail

Metadata summary

Number of items by collection, document type, and available metadata.

Doc. type	Subbasic	basic	advanced	Searchable full text	**Total**
Journal article (retrodigitised)					
Journal article (born digital)					
Book (retrodigitised)					
Book (born digital)					
...					
Other					
Total					

Notes: An *item* is a self-contained mathematical text which has been scientifically validated and formally published. A monograph is an item, not its chapters; a journal article is an item rather than a full journal issue. A proceedings volume or any collective book may be an item, as well as each of its individual contributions.

Contributed items must have a full text to be counted (EuDML is not a metadata-only project). Recent full texts need not be open access: give insight on moving wall policy and number of items protected by moving wall.

The metadata taxonomy has the following meaning:

- An item has **subbasic** metadata if it is not sufficient to reconstruct its basic bibliographic reference so as to locate it in a paper library, e.g., or to identify it without ambiguity. A typical

example of subbasic metadata is the pair: title/author, possibly with the mention of a source such as journal or a publisher.

- An item has **basic** metadata if the metadata identifies it unambiguously not only among the contributor's items, or even EuDML items, but among the whole literature. For a book, it means to have at least author names, title, publisher, edition (and hopefully the fact that it's a book). For a journal article, it means the journal title, volume, issue, page numbers or any other identifier system relevant for this journal (and hopefully the fact that it's a journal article). Basic metadata is the minimum requirement for setting up a reasonable browsing of the collections, as well as for matching a written reference with an item. Keywords or abstracts are appreciated components of basic metadata, although not mandatory to be qualified as such.
- **Advanced** metadata is a much wilder subject. Here, it means anything beyond basic. Reference lists, additional multilingual metadata (such as translated titles, keywords, abstracts), subject classifications, MathML formulae. Please specify.
- **Searchable full texts** can be extracted from a PDF, generated by unverified OCR, manually keyed, converted from TeX, be structured XML with MathML formulae. Please specify.

Expected updates

Please give insight on the expected number of new items contributed during the next

2 years, and the frequency of the updates.

Interoperability

The EuDML foreseen architecture is based on OAI-PMH harvesting of metadata (possibly OAI-ORE for harvesting various versions of full texts too). The EuDML consortium can help providers in setting up the necessary infrastructure: technical advice on setting up the server and preparing the metadata for export (WP3); conversion of metadata to EuDML format after harvest (WP5).

Do you have an OAI-PMH server serving your contributed collections' metadata?

- **If no: Do you have the capability of setting up one within the next 6 months?**
  Would you provide a set of XML files or a database dump to one of the EuDML partners thus serving as a broker for your metadata?
- **If yes: Does your OAI-PMH server only serve oai_dc metadata format?**

  Can you coerce it into obeying best practices suggested by project Euclid, SUB Göttingen and NUMDAM

  (http://projecteuclid.org/collection/euclid/documents/metadata/dml_dc.html)?

  Can you deliver any other metadata schema as soon as it is clearly defined?

Thank you!