

Survey of Existing Services in the Mathematical Digital Libraries and Repositories in the EuDML Project

Radoslav Pavlov, Desislava Paneva-Marinova, and Georgi Simeonov

Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria

{radko, dessi}@cc.bas.bg, gsimeonov@math.bas.bg

Abstract. This paper presents a survey of the existing services provided by the digital libraries and repositories on mathematics of the content provider partners in the EuDML project. The purpose is to support the development of the concepts, criteria and methods for the continuous evaluation of these and new relevant existing services. The work was concentrated on the classification of the relevant services in order to specify a common evaluating structure.

Keywords: Multimedia Digital Libraries, Evaluation of DL Services

1 Introduction

The EuDML project [1] 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 [5][6].

An important part the project is proper testing, assessment and evaluation against the initial proposal and the completing environment. The evaluation in EuDML will support the process of decision about future sustainability, the process of benchmarking during the development and also during the design the requirement for clear interfaces for adding new content repository in the future. In this sense we will evaluate mainly EuDML and its providers and also eventual competitors. For this purpose, it will be considered organization facet (i.e. organizational context and sustainability), services facet (i.e. the technical and functional characteristics of the services) and the content facet (its quantity and quality).

This paper presents a survey of the existing services provided by the environment of the EuDML content provider partners. The purpose is to support the development of the concepts, criteria and methods for the continuous evaluation of these and new relevant existing services. The work was concentrated on the classification of the relevant services in order to specify a common evaluating structure. The classification scheme follows frequently used services in digital libraries, archives, publishing systems and other content management systems. For each partner system's, defined

groups of services for readers, authors, administrators and services for interoperability have been surveyed.

2 Classification of Existing Services

The main scopes of the survey are services related to readers, authors, administrators and interoperability services (interoperability services and protocols). Readers services cover all related to web user interface navigation web feeds, cross references, Web 2.0 services (tagging, comments, ratings, reviews, bookmarks, share this), statistical reporting, email-alert and online subscriptions. Authors services are considered that they have all readers services and according functionalities, but with difference that authors can submit articles, track citations and etc. Services for administrators include all related to system maintenance like the most general tasks as management of users, groups and roles, metadata curation etc. Interoperability services are considered to the ways in which digital repositories and libraries work with other systems using common standards and protocols. Sometimes these interfaces are used directly by people (e.g. web user interfaces, web search engines or web feeds like RSS feeds) and sometimes they are used by machine-to-machine. For example OAI-PMH service has been built to interact with different repository implementations used to harvest (or collect) the metadata descriptions of the records in an archive so that services can be built using metadata from many archives.

The services provided in the EuDML partners systems could be grouped in the following common structure [7]:

Readers services (with or without required user registration/account)

- User interface (content presentation, content classification scheme, content retrieval, browse, navigation, simple/advanced search)
- Web feeds (RSS, Atom)
- Cross reference (citation linking mechanisms, link resolvers)
- Web 2.0 services (tagging, comments, ratings, reviews, bookmarks, share this, other Web 2.0)
- Statistical reporting (size, diversity, self-counting, count of total items, top downloads, top cited, collection/repository growth over time, etc.)
- User account (roles, groups)
- Email-alert
- Online subscriptions (free access/non-free paid, license model)

Authors' services

- User interface (content presentation, content classification scheme, content retrieval, browse, navigation, simple/advanced search)
- Web feeds (RSS, Atom)
- Cross reference (citation linking mechanisms, link resolvers)
- Web 2.0 services (tagging, comments, ratings, reviews, bookmarks, share this, other Web 2.0)

- Statistical reporting (size, diversity, self-counting, count of total items, top downloads, top cited, collection/repository growth over time, etc.)
- User account (roles, groups)
- Email-alert
- Online subscriptions (free access/non-free paid, license model)
- Online submissions (license model, copyright, ownership, terms of use, etc.)

Administrators' services

- Users, groups, roles management
- Metadata curation
- Customizable workflow
- Bulk import metadata
- Bulk export metadata

Interoperability services (machine-to-machine interoperability services and protocols)

- OAI-PMH
- OAI-ORE
- SWORD
- Harvesting
- Persistent identifiers
- Other

The survey is based on the frequently used services in digital libraries, archives, repositories and publishing systems. The purpose is to give us a clearer picture on of existing services and content providers in EuDML project.

The outcomes of the survey will facilitate the design decisions and recommendations for EuDML and EuDML partners.

3 Description of the Classified Services

3.1 Readers Services (User Interface Services)

Content presentation includes:

- File types of stored items (html, mathml, tex, ps, pdf, doc, images, etc).
- Service for online document format conversion – possibility for user to choose in which format to download document.
- Thumbnail (quick preview) - user can view content as a thumbnail (quick preview) in a web browser.
- Option to view full article through web browser without need to download as a file, etc.

Content classification scheme has a role in aiding information retrieval in a network environment, especially for providing browsing structures for subject-based information gateways on the Internet. Advantages of using classification schemes

include improved subject browsing facilities, potential multi-lingual access and improved interoperability with other services. A list of classification schemes and controlled vocabularies used in existing Internet services. Examples of classification schemes are Mathematics Subject Classification (MSC), DDC, UDC, Other.

Content retrieval includes actions such as export document in multiple formats and/or extract parts of a document (for example, extraction of only citations, references or figures from a particular article).

Browse and navigation covers browsing by author, subjects, year, title, collections, type of item (article, proceeding, book, etc) and other; content filtering (new, recent, key words, subject, similar, sort results), service for personalized seeking of similar articles/documents, etc.

Simple search represents searching by keywords, phrases; predefined search only on metadata field/s (title, abstract, etc.).

Advanced search: User can choose different scopes of search. Advanced search also provides any combination of searching at the same time by multiple selected scopes of search linked with conditional logical operators AND, OR, NOT. Advanced search may provide autocompletion of search terms; make suggestions with relevant keywords, phrases associated to the user search request.

Web feeds are used to publish frequently updated works such as blog entries, news headlines, etc. in a standardized format. An example of web feed format is RSS document (which is called a "web feed", or "channel") includes full or summarized text, plus metadata such as publishing dates and authorship. Web feeds benefit publishers by letting them syndicate content automatically. They benefit readers who want to subscribe to timely updates from favored websites or to aggregate feeds from many sites into one place. RSS feeds can be read using software called an "RSS reader", "feed reader", or "aggregator", which can be web-based, desktop-based, or mobile-device-based. A standardized XML file format allows the information to be published once and viewed by many different programs. The user subscribes to a feed by entering into the reader the feed's URI or by clicking a feed icon in a web browser that initiates the subscription process. The RSS reader checks the user's subscribed feeds regularly for new work, downloads any updates that it finds, and provides a user interface to monitor and read the feeds.

Cross reference (linking mechanisms, link resolvers): Some examples are Open URL linking, Link resolvers, Electronic resource integration, DOI, CrossRef, Handle.Net

Web 2.0 services: tagging, comments, ratings, reviews, bookmarks, share this, other.

Statistical reporting: size, diversity, self-counting, count of total items, top downloads, top cited, collection growth over time, etc.

User account (roles, groups): Here it should be listed and described any predefined users' roles and groups and the specific services for each of them.

Email-alert: Services related to the notification via e-mail

Online subscriptions (free access/non-free paid, license model): Terms and conditions. Definitions of offered services, access and use, policy, copyright, etc.

Open-Access License models: The more we understand about science and its complexities, the more important it is for scientific data to be shared openly. It's not useful to have ten different labs doing the same research and not sharing their results focusing efforts to expand the use of Creative Commons licenses to scientific and technical research [8].

License model applied by Public Library of Science (PLoS) is Creative Commons Attribution License (CCAL) [10] to all published works (see the human-readable summary or the full license legal code at [10]). Under the CCAL, authors retain ownership of the copyright for their article, but authors allow anyone to download, reuse, reprint, modify, distribute, and/or copy articles in PLoS journals, so long as the original authors and source are cited [9].

3.2 Authors Services

Online submissions (License model, copyright, ownership, terms of use, etc.): Terms and Conditions, Definitions, services offered, access and use

User interface, Web feeds, Cross references, Web 2.0 services, Statistical reporting, User account, Email-alert and Online subscriptions are same as readers.

3.3 Administrators Services

Users, groups, roles management: Describe existing types of users and roles.

Metadata curation describes existing metadata curation system and how it's implemented. The goal of the curation system is to provide a simple, extensible, way to manage routine content operations on a repository. Some examples are:

- ensure a given set of metadata fields are present in every item, or even that they have particular values
- profile a collection based on format types - good for identifying format migrations
- network service to enhance/replace/normalize an item's metadata or content
- ensure all items are readable and agree with the ingest values.

Customizable workflow: A workflow consists of a sequence of connected steps. It is a depiction of a sequence of operations, declared as work of a person, a group of persons, an organization of staff, or one or more simple or complex mechanisms. Workflow may be seen as any abstraction of real work. For example the workflow may consist of maintaining publications by importing metadata from other sources, and attaching full text where available. This minimizes the amount of manual form-filling needed. The interaction with the repository is limited to selecting which collection (if any) they want their work archived.

Bulk import metadata and Bulk export metadata: It's often more efficient to import and export at once large amount of data. Is it possible for example - export at once all collection, all content of repository software?

3.4 Interoperability Services

This section includes machine-to-machine interoperability services and protocols such as OAI-PMH, OAI-ORE, SWORD, Harvesting, and Persistent identifiers.

The classification scheme of existing services is based on frequently used services in digital libraries, archives, publishing systems and other content management systems. It's not obligatory and not expected for each content provider to have all of these services. For each partner system, we survey groups of services for readers, authors, administrators and interoperability. It will be collected data about maintenance of these services or similar. This survey will give a detail picture of the existing services during the first evaluation run. The outcomes of the survey will facilitate the design decisions and recommendations for EuDML system features and its functionality. Moreover, it could support the following groups of services with base priority for EuDML project [2][4]:

- *Services concerning to Interoperability and Integration* - describe the ways in which repositories work with other systems using common standards and protocols. Sometimes these interfaces are used directly by people (e.g. web user interfaces or RSS feeds) and sometimes they are used by machines (e.g. OAI-PMH and SWORD). Interfaces used by machines are sometimes referred to as m2m (machine-to-machine) interfaces.
- *Services supporting linking mechanism* - for effective use of distributed electronic resources in libraries. Some examples are Open URL linking, Link resolvers, Electronic resource integration, DOI, CrossRef, Handle.Net. Linking mechanism makes possible to build global digital libraries services and portals, because it provides unique item identifiers, persistent identifiers are used for citation management, etc.
- *Storage and long term preservation of digital information* – it concerns to using well known standards for metadata, storage data formats, etc. with provided support for a long time. Policies according to systems and software management, physical security, data security, data backups, disaster recovery, redundancy of data (multiple data duplication, digital archives, global web portals, providing content aggregation from various sources distributed over the Internet), etc.

4 Methodology

The classification scheme of existing services is based on the frequently used services in digital libraries, archives, publishing systems and other content management systems.

Generalized classification scheme [7] covers the services for readers, authors, administrators and interoperability services and is built on the base of modern content management systems with advanced, developed and rich functionalities that are presented in the following list:

- *arXiv.org* is an archive for electronic preprints of scientific papers

- *ACM Digital Library* - Full text of every article ever published by ACM and bibliographic citations from major publishers in computing.
- Springer Science or Springer is a global publishing company which publishes books, e-books and peer-reviewed journals in science. Springer also hosts a number of scientific databases, including SpringerLink, SpringerProtocols, and SpringerImages.
- *The Public Library of Science (PLOS)* - is a nonprofit organization of scientists and physicians committed to making the world's scientific and medical literature a public resource.
- *PLOS ONE* (accelerating the publication of peer-reviewed science) - An interactive open-access journal for the communication of all peer-reviewed scientific and medical research.

Readers services cover all related to web user interface navigation web feeds, cross references, Web 2.0 services (tagging, comments, ratings, reviews, bookmarks, share this), statistical reporting, email-alert and online subscriptions. Web 2.0 services especially web feeds could provide very useful additional mechanism for aggregating data. It gives possibility for users to personalize and subscribe for multiple feeds channels listed on centralized web portal. Other Web 2.0 service – social bookmarks can be integrated in EuDML concerning to book-shelves and collecting favourite links by users.

Authors services are considered that they have all readers services and according functionalities, but with difference that authors can submit articles, track citations, etc. These services are included because the most of observed systems have online manuscript submission system and other related services. Authors services in EuDML should be considered only for new digital born articles and for digital publishers.

Statistical reporting is useful not only for authors and readers, ratings, popularity and etc, but it can be useful in system interoperability for optimizing performance. For example metadata/content aggregators consume a lot of systems computation performance and network bandwidth when they retrieve content from large digital repositories. With statistical reporting can be determined and planned time schedules for aggregators when they should be used and how often. Time schedule for aggregators can be adaptive if there are monitoring services based on statistical reporting from content providers.

Services for administrators include all related to system maintenance like the most general tasks as management of users, groups and roles, metadata curation, etc.

The last group of services concerns to interoperability and integration. The interoperability and metadata aggregation are with primary priority according to EuDML project. It takes more effort to get support for machine-to-machine interfaces from all EuDML partners, because the most of them have no OAI-PMH or other protocol.

The collected data for this survey also includes technical specifications for content provider system software platforms, operating system, database, programming languages, additional functionalities, etc. The outcome from technical part of the survey could give general information based on used technologies for software sustainability, scalability and development.

In the Appendix A the service description tables of the Bulgarian contribution in the EuDML project - the Bulgarian Digital Mathematics Library [2][3], are presented. The data for the tables are collected by questionnaire or data collection sheet. Tables are based on classification scheme of existing services, frequently used in digital libraries/archives/publishing systems and other content management systems. Every EuDML content provider is filled this questionnaire. It's not obligatory and not expected for each content provider to have all of these services. For the system, we survey groups of services for readers, authors, administrators and interoperability. The outcomes of survey will identify the most used services by the content providers and therefore it facilitates choosing relevant services for EuDML project.

Acknowledgements. This work is supported by the international project CIP-ICT-PSP.2009.2.4 PROJECT „EuDML – European Digital Mathematical Library” (2010-2013) and the EU Competitiveness and Innovation Framework Programme, ICT Policy Support Programme.

References

1. European Digital Mathematical Library: <http://eudml.eu>
2. Dodunekov, S., Pavlov, R.: Digitization of Scientific and Cultural Heritage and the Bulgarian Digital Day Order”, National Workshop “Bulgarian Digital Day Order”, Sofia, Bulgaria (2011)
3. Pavlov, R.: Bulgarian Digital Mathematics Library – Short Presentation of the Project, National Workshop “Open Access to the Scientific Information, IMI-BAS, Sofia, Bulgaria (2010)
4. Pavlov, R.: No Royal Road but at Least a Gateway to the Mathematical Knowledge: The EuDML Project, Mathematics and Education in Mathematics, Proceedings of the Fortieth Jubilee Spring Conference of the Union of Bulgarian Mathematicians, pp. 70-79, Borovetz, Bulgaria (2011)
5. Pavlov R.: Save Past for the Future - Digital Presentation of Cultural and Science Heritage in Bulgaria, Jubilee Session "50 Years from the Creation of the First Computational Center in Bulgaria", Sofia, Bulgaria (2011)
6. Pavlov, R., Ikonov N.: Digital Libraries for Scientific and Cultural Content – Realization and Projects of IMI-BAS, International Conference “Steps, needed for the Integration of Bulgaria in the European Initiative Europeana”, Plovdiv, Bulgaria (2010)
7. Pavlov, R., Pelikan, P.: Report on evaluation of existing services and content providers (Deliverable 11.1, Public report), The European Digital Mathematics Library project (2011)
8. Science - Creative Commons <http://creativecommons.org/science>
9. PLoS ONE Open-Access License <http://www.plosone.org/static/license.action>
10. Creative Commons Attribution License (CCAL) <http://creativecommons.org/licenses/by/2.5/>

Appendix A

EuDML content provider: Institute of Mathematics and Informatics at Bulgarian Academy of Sciences

Provider's platform: Bulgarian Digital Mathematics Library (BulDML)

Web address: <http://sci-gems.math.bas.bg>

Table 1. Readers Services in BulDML

Readers Services							
User interface							
Content presentation	File types of stored items for presentation: pdf						
Content classification scheme	No						
Content retrieval	pdf, txt, every publication has exportable DC metadata						
Browse and navigation	Browse by community, collection, title, author, issue, year,						
Simple search	Yes						
Advanced search	Advanced search in metadata by any combination of searching at the same time by multiple selected scopes of search (collection, title, author, language, year, identifier), with conditional logical operators AND, OR, NOT.						
Web feeds	RSS			Atom			
	Yes [but disabled]			No			
Cross reference (<i>linking mechanism, link resolvers</i>)	Article itself referencing are handled by prefixes assigned and resolved by Hadle.NET system.						
Web 2.0 services	Tagging	Comments	Ratings	Reviews	Bookmarks	Share this	Other
	No	No	No	No	No	No	No
Statistical reporting	No						
User account (roles,	Yes, defined user groups: submitters, administrator						

groups)	
Email-alert (services related to the notification via e-mail)	Yes
Online subscriptions (free access/ non-free paid, license model)	Free access according to the policy of publishers. (all available content with free open access)

Table 2. Authors Services in BulDML

Authors Services	
Nothing special for authors	
Online submissions (License model, copyright, ownership, terms of use, etc.)	Online submission is available for authors according to the policy of publishers. Every document in the digital repository is provided with a flyleaf containing information about copyright and terms of use.

Table 3. Administrators Services in BulDML

Administrators Services	
Users, groups, roles management	Yes, management for users and groups
Metadata curation	Yes, (As of release 1.7, DSpace supports Curation System and can running curation tasks)
Customizable workflow	Yes
Bulk import metadata	Yes
Bulk export metadata	Yes

Table 4. Interoperability Services in BulDML

Interoperability Services (machine-to-machine interoperability services and protocols)	
OAI-PMH	Yes
OAI-ORE	No
SWORD	Yes
Harvesting	No
Persistent identifiers	Yes (prefixes assigned and resolved by Handle.NET System)
Others	No