Vittorio Coti Zelati
bdim: the Italian Digital Mathematical Library

In: Petr Sojka (ed.): Towards a Digital Mathematics Library. Paris, France, July 7-8th, 2010. Masaryk University Press, Brno, Czech Republic, 2010. pp. 79--81.

Persistent URL: http://dml.cz/dmlcz/702576

bdim: the Italian Digital Mathematical Library

Vittorio Coti Zelati

Dipartimento di Matematica e Applicazioni
Università degli Studi di Napoli "Federico II"
via Cintia, M.S. Angelo
80126 Napoli, Italy

Abstract. We present bdim (Bibliteca Digitale Italiana di Matematica) the Italian project of math digitization. The project has been started by SIMAI (Società 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 Attività Culturali and with the help of Numdam. At the moment bdim consists of approximately 1,300 articles, 11,000 pages (articles from Bollettino Unione Matematica Italiana, 1946–1967).

1 Italian Math Journals

In Italy there are many math journals published by Mathematics Departments or by Scholarly Societies. In the last few years many of them have started distributing the full-text online and some of them have decided to switch to a commercial editor. A non exhaustive list of journals and of their present situation is the following:

- Journals that have decided to go with a commercial editor: *Annali di Matematica Pura ed Applicata* (Springer since 2001, online on SpringerLink), *Rendiconti Lincei – Matematica e Applicazioni* (EMS Publishing House since 2005, online with the publisher starting from the 2005 volume), *Ricerche di Matematica* (Springer since 2006, online on SpringerLink starting from the 2006 volume), *Annali dell'Università di Ferrara* (Springer since 2006, online on SpringerLink), *Rendiconti del Circolo Matematico di Palermo* (Springer since 2008, online on SpringerLink)
- Journals that have been digitized by Numdam: notably *Annali della Scuola Normale Superiore di Pisa. Classe di Scienze* and *Rendiconti del Seminario Matematico della Università di Padova*.
- Journals which have not been (or have been only partially) digitized: *Bollettino dell'Unione Matematica Italiana*, *Note di Matematica* (Lecce, online in Lecce), *Rendiconti di Trieste* (on-line in Trieste), *Rendiconti di Matematica e delle sue Applicazioni* (Roma, partly on-line in Rome), *Atti del Seminario Matematico e Fisico dell'Università di Modena* (Modena), *Istituto Lombardo. Accademia di Scienze e Lettere. Rendiconti. Scienze Matematiche e Applicazioni* (Milano), *Le Matematiche* (Catania), *Rivista di Matematica della Università di Parma* (Parma),

Petr Sojka (editor): DML 2010, Towards a Digital Mathematics Library, pp. 79–81.
 ISBN 978-80-210-5242-0

Università e Politecnico di Torino. Seminario Matematico. Rendiconti (Torino, partly on-line).

There are also Journals which are no more published, like the *Giornale di Matematiche di Battaglini*.

We do not have precise figures, but for sure more than 300,000 pages are still to be digitized (about 100,000 for the Bollettino dell'Unione Matematica Italiana).

Of the different categories, only the journals which have decided to go with Numdam are fully integrated with DML. Two of the Journals now with a commercial editor (*Rendiconti Lincei* and *Ricerche di Matematica*) have not been fully digitized, and should join our project.

2 The bdim Project

The bdim project has been started by SIMAI (Società Italiana di Matematica Applicata e Industriale) and UMI (Unione Matematica Italiana) in connection with the international effort to create DML, the Digital Mathematical Library and in collaboration with the BDI (Biblioteca Digitale Italiana, the Italian Digital Library).

Our goal is to search for funding, digitize and provide to the many Italian math journals a commun repository, a better international visibility and an interface to DML and BDI.

The initiative has received a grant of 15,000 Euro from the Italian Government to define the standards, to digitize some material and to implement the repository and the web interface needed for the dissemination of the material.

At the moment, we have digitized the *Bollettino dell'Unione Matematica Italiana Serie III* (published in the period 1946–1967, consisting of 22 volumes, 83 issues, 1,358 articles, 11,390 pages) and we are testing the repository. The test repository is accessible at the address `http://bdim.dma.unina.it/` (it will change). We have also implemented an OAI-PMH server, compatible with the Mini-DML standard, accessible at the address `http://bdim.dma.unina.it:8080/oaiprovider/?verb=Identify` (also this address will change).

We are now acquiring also the *Bollettino dell'Unione Matematica Italiana Serie VIII*, period 1998–2007. Meanwhile we are extending the initiative to other italian journals and we are seeking for the necessary financial support.

3 The Implementation

In implementing our project we have tried to follow the example of Numdam, from whose staff we have had a lot help and advice. We have also adhered to the standards suggested by the DML project and to those required by the Biblioteca Digitale Italiana, in particular the MAG standards (Metadati Amministrativi e Gestionali, `http://www.iccu.sbn.it/genera.jsp?id=267`) required by the

Istituto Centrale per il Catalogo Unico delle Biblioteche Italiane e per le Informazioni Bibliografiche `http://www.iccu.sbn.it/`.

Each page has been digitized at the resolution of 600 dpi (300 dpi for pictures, photos, etc). To each page has been associated the text obtained with OCR. Each issue has then been segmented into articles, and and a PDF and a DjVu file have been produced for each article.

We have decided to use a Fedora repository (`http://www.fedora-commons.org/`) to manage, organize and preserve our material. Fedora is a flexible, well tested project, platform independent and well suited for our needs. From the overview of Fedora:

> In a Fedora repository, all content is managed as data objects, each of which is composed of components ("datastreams") that contain either the content or metadata about it. Each datastream can be either managed directly by the repository or left in an external, web-accessible location to be delivered through the repository as needed. A data object can have any number of data and metadata components, mixing the managed and external datastreams in any pattern desired.
> Each object can assert relationships to any number of other objects, providing a way to represent complex information as a web of significant meaningful entities without restricting the parts to a single context.

Our (Fedora) repository has been organized as follows: we have identified as *digital object* each journal, each volume, each issue and each article. Each digital object has at least one datastream: an XML file which contains both the administrative and gestional information on the object as well as the bibliographical information on the object itself. Each digital object might have some other datastream: for example the articles objects have the corresponding PDF, DjVu and OCR files, the issues the tiff files of all the pages making up the issue.

The Fedora repository (`http://bdim.dma.unina.it:8080/fedora/`, not accessible from the web) runs on tomcat web server. The web interface (`http://bdim.dma.unina.it/`) to the repository runs on a different web server (at the moment on the same machine), is written in PHP, obtains the data from the repository and (modulo an XSLT transformation) disseminate it. For example, to build the page for an article, or an issue of a journal, the web server "ask" to the repository for the relevant XML file, and builds the web page from there. The search engine (not yet implemented in the test repository where a simpler search engine is running) is based on a service of Fedora (GSearch) which selectively harvests content and metadata from objects and indexes them. Also the OAI server `http://bdim.dma.unina.it:8080/oaiprovider`, which is another of Fedora's services, harvests the data from the Fedora server and disseminate it.