

# A user-centred portal for search and retrieval of open-access Italian scholarly literature: the PLEIADI project

*Ugo Contino\**, *Paola Gargiulo\**, *Susanna Mornati\*\**, *Zeno Tajoli\*\**

\* CASPUR

Via dei Tizii 6 b, 00185 Roma Italy

[contino@caspur.it](mailto:contino@caspur.it)

[gargiulo@caspur.it](mailto:gargiulo@caspur.it)

\*\* CILEA

Via R. Sanzio 4, 20090 Segrate (MI)- Italy

[mornati@cilea.it](mailto:mornati@cilea.it)

[tajoli@cilea.it](mailto:tajoli@cilea.it)

## Abstract

The PLEIADI Project (acronym for “**P**ortale per la **L**etteratura scientifica **E**lettronica **I**taliana su **A**rchivi aperti e **D**epositi **I**stituzionali”), a portal for Italian scholarly e-literature in open archives and institutional repositories) was born within the framework of a collaboration between two major Italian university consortia, CASPUR and CILEA, as part of a project called AEPIC (<http://www.aepic.it>). PLEIADI’s goal is to promote national awareness in the Open Access scenario and provide centralized access for scholarly literature archived in Italian institutional repositories (or “data providers” in the OAI - Open Archives Initiative architecture). PLEIADI also aims at providing end users (mainly the academic and research community) with a personalized environment, user profiling, an area to save searches and articles, e-mail alerts. The portal enhances users’ awareness on Open Access issues through several information services (news, RSS, forum), and promotes self-archiving of scholarly articles in institutional repositories. The service, available at the URL <http://www.openarchives.it/pleiadi/>, was launched in November 2004, during a national event held in Messina, which marked the “official starting point” for the Open Access movement in the Italian academic community.

## 1 Introduction

PLEIADI’s initiative was born in 2003 [1] within the context of the so-called “scientific communication crisis”. The information and communication technology progress, which constantly delivers new, powerful, and customizable end-user services, was expected to boost the rate of diffusion of scholarly electronic publications, and to lower their costs. But commercial publishers still set strong limitations on the circulation of scientific articles, both in the printed and in the electronic format. Several reasons account for this, the main one being economical. Universities and research centres cannot afford the rising costs that maintaining commercial subscriptions for all scholarly journals would put on their budgets. This constitutes a true paradox, since those who most suffer from these access limitations are the same authors that are responsible for content generation of scholarly literature. Moreover, most traditional publishers have not been able to introduce major changes in formats and media, nor to reduce the latency between the availability of a draft and the actual publication of an article. These limitations result in communication delays and serious loss of potential impact, that is the main interest of scientific communities.

In this critical situation we assisted to the rise of several initiatives in favour of Open Access (OA in the remainder of this contribution) to scholarly literature [2], promoting both open-access journals and self-archiving strategies. The former means that readers do not have to pay to access published articles, the latter implies that authors deposit an electronic copy of their own scholarly publications, together with simple descriptive data (metadata), in digital repositories that are freely and openly accessible on the Internet. These repositories support the OAI-PMH (Open Archives Initiative - Protocol for Metadata Harvesting) [4], the main outcome of the Open Archives Initiative [4], as a means of exposing metadata to federated search services. In the OAI architecture, the protocol provides an application-independent interoperability framework based on *metadata harvesting* from compliant repositories (*Data Providers*). *Service Providers* use metadata harvested via the OAI-PMH as a basis for building value-added services, including centralized search facilities.

## 2 PLEIADI – A national service

The PLEIADI Project (acronym for “Portale per la Letteratura scientifica Elettronica Italiana su Archivi aperti e Depositi Istituzionali”, a portal for Italian scholarly e-literature in open archives and institutional repositories) [5] originated from the collaboration between two major Italian university supercomputing consortia, CASPUR and CILEA [6], within the framework of the AEPIC project [7]. PLEIADI aims to create a common infrastructure where metadata about open access literature, harvested from Italian open archives (mainly institutional, and discipline based where appropriate), can be searched and retrieved centrally. Another goal is to enhance users’ awareness on Open Access issues. Through PLEIADI, scholars are encouraged to make their articles available through institutional repositories, librarians, administrators and IT people are provided with useful information for OA management, policy makers can measure the effects of the Open Access paradigm in the national context.

PLEIADI Portal for the Italian Electronic Literature in Open and Institutional Archives

**Login**  
 Username:   
 Password:   
  
 Lost Password?  
 Register now!

**Main Menu**  
 Home  
 News  
 RSS: News from OA world  
 Forum  
 Search in archives  
 Web Resources  
 PLEIADI Project  
 About us

Select your language

Powered by

**OpenArchives Search**  
   
 More searching options

**PLEIADI project manifesto**  
 The PLEIADI Project (acronym for "Portale per la Letteratura scientifica Elettronica Italiana su Archivi aperti e Depositi Istituzionali", a portal for Italian scholarly e-literature in open archives and institutional repositories) originated from the collaboration between two major Italian university consortia, CASPUR and CILEA, within the framework of the AEPIC project. [\[more...\]](#)

**Recent News**

- PLEIADI at ELPUB 2005 (2005/5/31)
- METALIS, a service for LIS open access literature (2005/4/28)
- Webcast: Ownership and Access in Scholarly Publishing (2005/4/18)
- OA policy implementation and Southampton outcomes (2005/4/18)
- OpenURL standard, NISO Z39.88 (2005/4/18)
- BBC Creative Archive (2005/4/18)
- Open Culture: accessing and sharing Knowledge, Scholarly Production and Education in the Digital Age, Milano June 27-29 2005 (2005/4/6)
- Columbia Univ. resolution on open access (2005/4/5)
- Stern against Open Access (2005/4/4)
- Automatic metadata generation (2005/4/4)

**Recent Topics**

| Forum  | Topic                      | Replies | Views | Last Post               |
|--------|----------------------------|---------|-------|-------------------------|
| OAI.it | Open Journal System (OJS)  | 0       | 164   | 2005/5/30 16:55 Bollini |
| OAI.it | Novità su sw Eprints       | 2       | 853   | 2005/5/19 9:21 tajoli   |
| OAI.it | MyOPIA e DC                | 0       | 352   | 2005/5/5 18:57 dascoli  |
| OAI.it | Sistemi di classificazione | 3       | 388   | 2005/5/5 16:24 Mornati  |
| OAI.it | IEEE                       | 0       | 507   | 2005/4/17 21:32 dascoli |

[Visit Forums](#)

SOME RIGHTS RESERVED

Fig.1 - PLEIADI home page

PLEIADI started as a demonstration project, to explore interoperability issues among repositories, and it became a service, to implement common search across repositories. It increases the visibility and impact of each data provider, enhancing awareness and dissemination, and offers a user-friendly interface, based on a personalized environment (see figure 1). Information services are tailored for the Italian OA community (Forum, News, RSS [8]) and alerting functionalities (e.g. recent news and forum postings) are also available for registered users. Registration is free and it is not required to access the search module, which has been specifically designed to allow retrieval of centrally-indexed metadata. Other services, such as individual search history and article storage, personalized alerting on newly harvested articles, statistical information on search results and access to documents, export facilities to build individual and/or institutional bibliographies or web pages for discovery or assessment, will be implemented in the near future.

PLEIADI's other goal is even more crucial to achieve. Promotion of awareness of OA culture in Italy is a task that CILEA and CASPUR are pursuing in the interest of local academic and research communities, particularly through advocacy initiatives and PLEIADI's services. Librarians and ICT experts are also addressed, since they are usually in charge of providing information services to scholars. Only by participation in the international OA movement can Italy maintain its visibility and keep a high profile in the research world.

### 3 PLEIADI – The Portal layer

PLEIADI's platform is built on a two-level architecture. Its visible layer is a portal, Behind the portal there is a cluster of *service providers*.

The portal, developed using XOOPS [9], an *open source* framework for *content management oriented* systems, has a modular architecture, constituted by several functional blocks that can be grouped as follows:

- a ***user area***, encompassing users' front-end, multi-language support and authentication/authorization system. It is a personalized environment where the user creates its own profile, activates different features from language selection to alerting services related to several information areas, and soon search history and article storage;
- an ***information area***, based on News, Forum and RSS blocks, together with a rich and constantly updated area devoted to commented links to OA resources on the web. A document area is also in construction, in order to collect and disseminate OA related materials;
- a ***search & retrieval area***, with two access possibilities: a simple search box, where users are asked to insert any search terms, and a more sophisticated one, where search filters are available to limit the search to single archives, to the year of publication, to the material type or the document classification. Sorting of results can also be changed by the user according different criteria. The connection between the portal front-end and the service providers back-end is based on the Z39.50 protocol [10];
- a ***management area***, where an intuitive GUI (Graphical User Interface), based on a *content management engine*, allows the administrators to manage both content and look of the web site;

- a *log & stats area*, tracking global statistical data (search activities, forum topics accesses, news readings, links rating, etc.). This area is partly public and partly available to administrators at the moment, while it is still under development;

XOOPS has been chosen for PLEIADI's portal implementation for several reasons, the main ones being:

- graphical and user-friendly portal management functions;
- dynamic generation of web pages thanks to a database back-end for contents, which is based on MySQL;
- availability of an open-source generator of style-sheet templates, SMARTY [11];
- PHP-HTML code independency.

Different modules, constituted by PHP scripts [12], are at the basis of each portal section; these modules, made available by the XOOPS developers' community, have been partially modified in order to meet accessibility standards and fulfil project requirements. All portal contents are stored in a MySQL database [13]; this feature, together with the usage of CSS (Cascading Style Sheets) [14], allows a complete code decoupling from the HTML layer. An Apache WEB server installation [15] completes the open source environment adopted for PLEIADI's design.

#### 4 PLEIADI – Service provider features

In figure 2, PLEIADI is represented by a UML (Unified Modelling Language) schema. The back-end role for service delivery is carried out by a system of service providers (SP), a functional side in the above-mentioned OAI architecture.

The system [16] performs several functions:

- harvesting of metadata from Italian open archives via the OAI-PMH protocol;
- filtering and *harmonization* of harvested metadata through crosswalks (i.e. XML tag mapping), including re-classification, in order to uniform them and provide uniform values for a consistent retrieval;
- centralized metadata indexing and caching in a local collective database;
- Z39.50 and HTTP interfaces to expose metadata to the portal search module and to export them to any other system for federated search via the Internet, including web crawlers;
- OPAC and OpenURL functions that at the moment are not being used for PLEIADI, but have been developed to exploit the SP system for other services of public purpose [17].

Most service provider modules used here are written in Perl or PHP scripting languages, adopting and modifying open source software and free code libraries available on Internet, such as the YAZ functions library [18], developed for PHP, on which the Z39.50 communication channel is based. On top of this infrastructure some original code development was required to create predefined query templates and parsing of query results, filtering and ordering functionalities, in order to get more benefits from the harvesting subsystem.

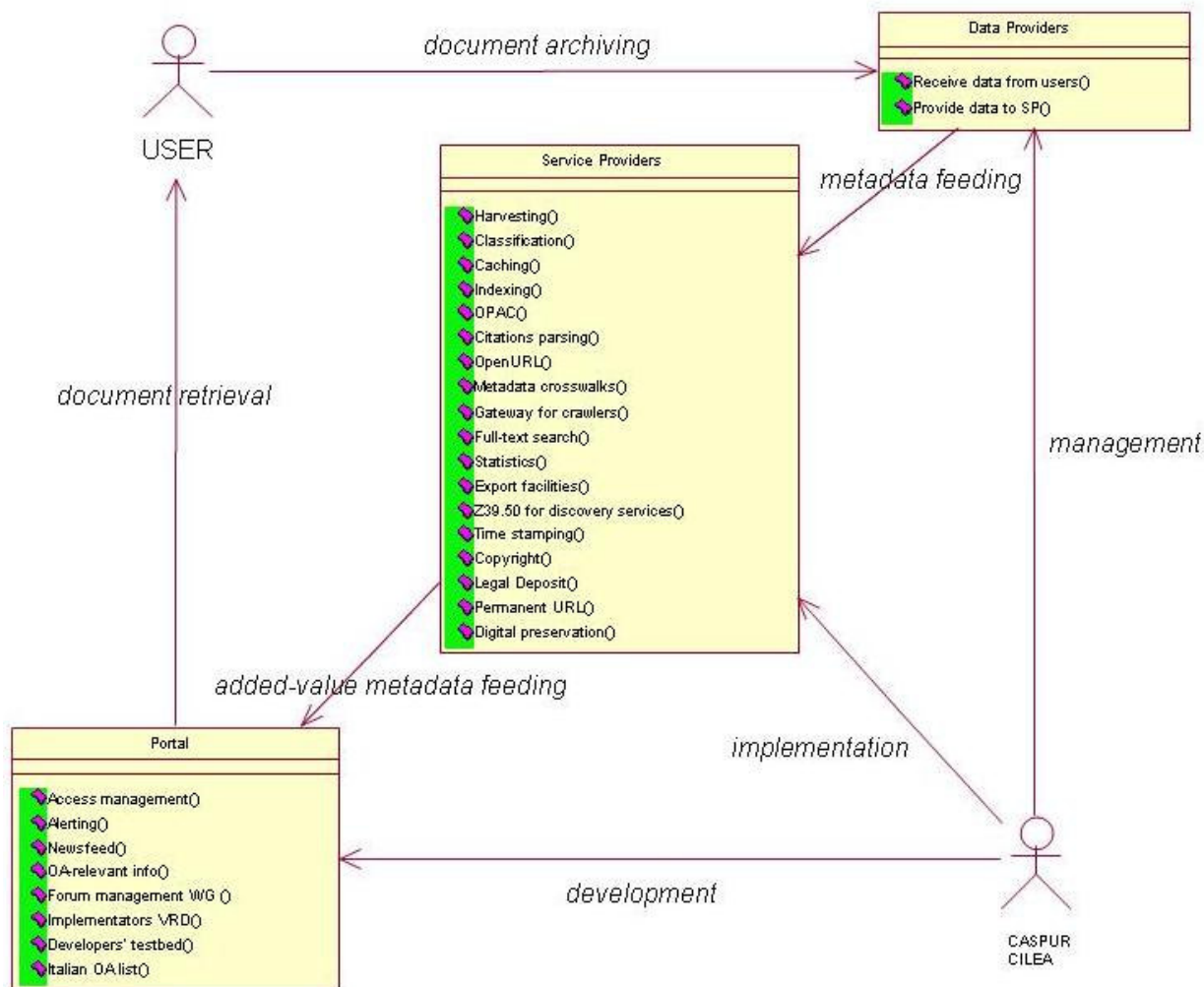


Fig.2 – UML schema of PLEIADI

Metadata harvesting has been implemented through Celestial [19], which was partially modified to handle selective harvesting by sets. This enables PLEIADI to select Italian literature both from Italian institutional repositories and from some international disciplinary open archives. Another noteworthy new functionality is a graphical interface to manage the addition of new repositories as they are notified to the service. Eight repositories (mainly institutional) are currently harvested from PLEIADI, providing more than 2,000 document descriptions (May 2005) in the central database.

Harvested metadata are expressed in the “oai\_dc” format [20], based on the unqualified Dublin Core [21]. Crosswalks are being written for each newly harvested repository, allowing conversion of different metadata values into a common format. Material types and classification schemas are typically different in each archive, and an accurate analysis and mapping activity is necessary to provide quality to a common search interface.

Cheshire II [22] has been chosen for indexing. It embeds two export interfaces, via Z39.50 and HTTP. Current PLEIADI’s design makes use of the Z39.50 channel; search criteria are included in the Z39.50 XML *payload*. Search results are passed to a PHP module in the portal system, built *ex-novo* for this project making use of DOM libraries [23]. The HTTP channel can be useful for further project developments.

## 5 Conclusions

PLEIADI was launched in November 2004, during a national conference [24] held in Messina, when more than 30 universities declared their support to the Berlin Declaration [25]. This event marked the “official starting point” for the OA movement in the Italian academic community.

Since then, the number of PLEIADI’s registered users has been constantly increasing, and recently crossed the threshold of 100 accounts. Forum area encompasses a number of different topics, and more than 50 items have been posted in the news area. Particularly rich is the *Web Resources* area, where about 180 commented links, are provided. All links are grouped in 19 categories, ranging from Italian advocacy to OA projects, from bibliographies to software and standards.

In the same period, the growth rate of Italian new OAI-compliant archives is also remarkable, and PLEIADI is planning to extend its harvesting activities to them as soon as they reach a significant amount of full-text, open-access research documents [26].

Even if PLEIADI is constantly growing in terms of users and services, much work is still ahead. Centralized download statistics, citation analysis, full-text search, copyright management, legal deposit, permanent URL, and digital preservation are foreseen, but can only be implemented when data providers support a common format to express necessary data and to export them via the OAI protocol. Particularly critical in terms of users’ expectations are several other user-related services, such as an individual search history and bibliographical storage space. With the growth of the number of available repositories also a centralized alerting system across repositories on newly deposited items will become crucial, and more crosswalks will be necessary. Last, but not least, statistical data management will provide useful information to both PLEIADI’s managers and users. These issues are being addressed by CASPUR and CILEA in a collaborative framework that will produce further services for the Italian scholarly community.

## References

All cited links have been checked on June 10<sup>th</sup> 2005.

- [1] Pleiadi – Open Archives Platform – Project Plan (in Italian); [http://www.aepic.it/documenti/PLEIADI\\_project\\_plan\\_ita\\_1\\_4.pdf](http://www.aepic.it/documenti/PLEIADI_project_plan_ita_1_4.pdf)
- [2] Definition provided by the *Budapest Open Access Initiative* (BOAI), <http://www.soros.org/openaccess>
- [3] OAI- PMH, <http://www.openarchives.org/OAI/openarchivesprotocol>
- [4] Open Archives Initiative (OAI), <http://www.openarchives.org/>
- [5] PLEIADI, [http://www.openarchives.it/pleiadi/index.php?sel\\_lang=english](http://www.openarchives.it/pleiadi/index.php?sel_lang=english)
- [6] CILEA: Consorzio Interuniversitario Lombardo per l’Elaborazione Automatica (<http://www.cilea.it>); CASPUR: Consorzio interuniversitario per le Applicazioni di Supercalcolo per Università e Ricerca (<http://www.caspur.it/en/>)
- [7] AEPIC (Academic E-Publishing infrastructures at CILEA), <http://www.aepic.it>, is a service that provides OA advocacy, training and consultancy for OA issues, installation and configuration of open archives and e-publishing applications for institutions and research groups.
- [8] *Really Simple Syndication* or alternatively *Rich Site Summary* – see <http://www-106.ibm.com/developerworks/library/w-rss.html> for an introduction to RSS technology; PLEIADI’s RSS

area is based at present on the weblog *Open Access News* by Peter Suber,  
<http://www.earlham.edu/~peters/fos/fosblog.html>

- [9] XOOOPS, <http://www.xoops.org/>
- [10] ANSI-NISO Standard; ref. <http://www.niso.org/standards/resources/Z39-50-2003.pdf>
- [11] SMARTY, <http://smarty.php.net/>
- [12] PHP, <http://www.php.net/>
- [13] MySQL, <http://www.mysql.com/>
- [14] See CSS *standard* description on <http://www.w3.org/Style/CSS/>
- [15] Apache, <http://www.apache.org/>
- [16] The whole system architecture is similar to the one used by the ePrints UK Project, <http://www.rdn.ac.uk/projects/eprints-uk/>. We are grateful to UKOLN and RDN for support in implementing this solution.
- [17] e.g. METALIS, <http://metallic.cilea.it/>, a service provider for Library and Information Science OA literature, also developed by CILEA in the AEPIC's framework.
- [18] YAZ, <http://zoom.z3950.org/>
- [19] Celestial, <http://celestial.eprints.org>
- [20] OAI Dublin Core: [http://www.openarchives.org/OAI/2.0/oai\\_dc.xsd](http://www.openarchives.org/OAI/2.0/oai_dc.xsd)
- [21] DCMI (2004), *Dublin Core Metadata Element Set, Version 1.1: Reference Description*: <http://dublincore.org/documents/dces/>
- [22] Cheshire II Project, <http://cheshire.berkeley.edu/>
- [23] DOM, <http://it2.php.net/manual/en/ref.domxml.php>
- [24] *Gli atenei italiani per l'open access; Workshop* – University of Messina; November 4<sup>th</sup>-5<sup>th</sup> 2005; <http://www.aepic.it/conf/index.php?cf=1>
- [25] *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*, <http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html>
- [26] A list of notified Italian open archives is available <http://www.openarchives.it/>