D-Lib Center – IST-2001-32587
Digital Library Competence Center

Digital Library Based Collaboration Testbed – Report: Deliverable 3.1.1

Deliverable Type: Public Report
Number: D3.1.1
Contractual Date of Delivery: 30 November 2003
Nature: Public

Task WP3

Name of responsible: Costantino Thanos (ISTI-CNR)

Authors: Umberto Straccia, Leonardo Candela, ISTI-CNR

Contact info: umberto.straccia@isti.cnr.it

Abstract

This workpackage had the aim to present Digital Library technologies which enable collaboration between Digital Library patrons. The complete documentation has been made available to the public through the D-Lib Center web site (http://dlibcenter.iei.pi.cnr.it/).

Executive Summary

This work-package had the aim to present digital library technologies which enable collaboration between digital library patrons.

The D-Lib Center established, operated and managed a digital library based collaboration environment. This environment is built on top of the Open Archive Initiative (OAI) which develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content. In particular, the D-Lib Center made available the Cyclades digital library system (http://www.ercim.org/cyclades), developed within a project funded by the IST Programme of the European Commission (5^{th} FP). The CYCLADES Service Environment provides two kinds of OAI compliant functionality: (i) functionality which supports a community/group of scholars while using the OA content as a basis for collaborative work (community functionality), and (ii) functionality which supports a single user/scholar when interacting with the OA environment (user functionality).

A course on open access to digital libraries was organised and held at the D-Lib Center laboratory on the following date:

20-21 November 2003

Teaching Staff:
Umberto Straccia, ISTI - CNR

Technical Staff:
Leonardo Candela, ISTI – CNR

Table of Contents

1. Course overview

The course provided two days activities at the D-Lib Center, organized as follows:

- What is a Digital Library
  - History,and basic notions about digital libraries are given
- Building a Digital Library
  - Issues related to the creation of a digital library are presented
- The Open Archive Initiative
  - Open Archive Initiative is described and related tools are described
- The Collaborative Working Environment CYCLADES
  - Fundamentals of collaborative work in CYCLADES are presented
- Experiencing CYCLADES
  - Features of the BSCW system (on top of which CYCLADES is built)
  - Features of the CYCALDES system are demonstrated, not covered by the BSCW system
- An example of collaborative work: managing the CYCLADES project
  - It is shown how the CYCLADES project has been managed using the collaborative working environment BSCW
- Discussions
  - Discussion forum
- Questionnaire

The demonstration and the discussion forum activities were based on the CYCLADES system. The documentation distributed at the course and available for download is composed of: 1) Copy of the slides used during the course. 2) User manual of the CYCLADES system.

2. Tutorial

This section provides some more details on the different topics presented during the course.

2.1. What is a Digital Library

This part of the course began with a definition of what is a Digital Library and made a comparison between traditional and new generation digital libraries. The differences and the similarities in terms of data types managed and services offered were illustrated. In addition, the main application areas of digital libraries were analyzed. This introduction continued with a detailed list of new digital objects and multilingual documents which may be maintained in the same DL and it illustrated the fact that new document types impose a re-thinking of the “traditional” library services. This part ended showing an example of a collaborative environment

2.2. Building a Digital Library

This part of the course aimed at showing how to design and build a digital library giving an overview of the E-Prints and DSpace software and illustrating three application scenarios.

2.3. The Open Archive Initiative

In this part of the course the Open Archive Initiative was explained in detail. First, the Protocol for Metadata Harvesting (OAI-PMH), the OAI Major Accomplishment, was illustrated. Since the protocol for Metadata Harvesting is an almost stateless request/response protocol and responses are well-formed XML documents, it was also given a brief introduction to XML.

2.4. The Collaborative Working Environment CYCLADES

A personalized collaborative Digital Library environment was introduced and explained in detail in this part of the course. CYCLADES is in fact a personalized collaborative Digital Library environment where users may organise the information space according to their own subjective view, may become aware of each other, may exchange information and knowledge with each other, may share opinions, may build communities and get recommendations based on preference patterns of the users.
The following aspects of CYCLADES were analysed: its objectives, functionality and architecture.

2.5. Experiencing CYCLADES

Features of the BSCW system (on top of which CYCLADES is built) and features of the CYCLADES system, not covered by the BSCW, were demonstrated in this part of the course and it was given an example of collaborative work showing how the CYCLADES project had been managed using the BSCW working environment.

2.6. Discussion Forum and questionnaire

At the end of the two-day course it was held a discussion forum to talk about the feasibility of the CYCLADES system and a questionnaire asking for the student's interest and satisfaction on the course was distributed and filled out.

3. Course organisation

One course was organised and held on November 20-21, 2003 at ISTI-CNR, in Pisa. The equipment of the D-Lib Center Lab, composed of a server running the CYCLADES system, and five workstations used as clients, was used for the course.
The course was advertised in a large number of specialised mailing lists (see Appendix C) and through direct mailing contacts. A flyer (see Annex B) advertising the course was distributed to all major international conferences in the field.
The course on "D.L. Based Collaboration" was attended by 15 persons; 10 filled in the questionnaire designed to evaluate the course. Nine participants were Librarians coming from Universities/Research Institutions and only few of them had a background in computer science

4. Appendix A

This appendix provides an overview of the CYCLADES project and of the results obtained.

A.1 Introduction to the CYCLADES System

CYCLADES is a system, which combines several technologies from the Information Retrieval and Digital Library area, where users and user communities may deal with a quite large set of heterogeneous digital archives. CYCLADES provides a highly personalized environment where not only users may organize (and search into) the information space according to their individual taste and use, but, and more importantly, also provides advanced features of collaborative work among the users. It is up then to the system to discover interesting properties about the users' interests, relationship between users and user communities, as well as meaningful events that happen in user shared workspaces, and finally to notify the involved users according to their own preferences.

It is widely recognized that *Digital Libraries* (DLs) will play an important role in the next future not merely in terms of the 'controlled' digital information they allow access to, but in terms of the *services* they provide to the information society at large. Informally, DLs can be defined as consisting of collections of information (usually, heterogeneous in content and format), which have associated services delivered to users and user communities using a variety of technologies. The services offered on such information can be varied, ranging from content operations to rights management and can be offered to *individuals* as well as to *user communities*. Indeed, an essential technology component of DLs is that they are networked, meaning that access is increasingly becoming *shared* and *collaborative*.

Even though DLs have evolved rapidly over the past decade, typically, DLs still are limited to provide a search facility to the digital society at large. Indeed, they are oriented towards a generic user, as they answer queries crudely rather than, *e.g.* learn the long-term requirements of a specific user. In practice, users use the same information resource over and over and would benefit from customization: the time consuming effort that the user put in searching documents and possibly downloading them from the DL is often forgotten and lost. This requires a repetition of the manual labour in searching and browsing to find the documents just like the first time. As DLs will become more commonplace and the range of information they provide and the services upon increases, users' expectations will increase and users are expecting more and more sophisticated services from their DLs. A 'quick and dirty search' facility is normally an integral part of any digital library, but users' frustrations increase as their demands become more complex and as the volume of information managed by digital libraries increases. There is a need for DLs to move from being *passive* with little adaptation to their users, to being more *proactive* and *personalized* in offering and tailoring information for individual users. Personalization can be defined as the way in which information and services can be tailored in a specific way to match the unique and specific needs of an individual user or a community of users. This is achieved by adapting the presentation and/or the services presented to the user by taking into account the user's task, background, history, device, information needs, location, etc., essentially the user's context. Personalization can be user-driven which involves a user directly invoking and supporting the personalization process by

providing explicit input, or personalization can be completely automatic, where the system observes some user activity and identifies the input used to tailor some aspect of the system in a personalized way. These two examples of user-driven and automatic personalization are at the extreme ends of the spectrum and many personalization tools will have elements of both approaches.
Nowadays, in several DLs some personalization functionalities are provided. Mainly they fall into the category of personalized *alerting services*, *i.e.* services that notify a user (usually, by sending an e-mail), with a list of references to newly available documents in the DLs and deemed as relevant to some of the (manually) user specified topics of interests. Some other DLs, in addition, support the users in being able to organize their information space they are accessing to according to *their own subjective perspective*. This is important as not necessarily all the information provided by a DL may be of interest to an user, but just some 'slices' of it. Users and communities of users might well profit from being able to organize the information space in a personalized fashion both int terms of restricting the information space in which to search into as well as in terms of organizing it not necessarily in the way a the DL manager thought would be well-suited for anyone.

We present the CYCLADES system and stress its 'personalization' and alerting features. A major distinction of CYCLADES is the fact that it envisages a DL not only as information space in which individual users may search for and organize the information provided by a DL, but also as a *collaborative meeting place* of people sharing common interests. Indeed, DLs may be viewed as a *common working place* where users may become aware of each other (indeed the system may find out interesting relationships both between users and/or between communities of users and produce the appropriate recommendations), open communication channels, and exchange information and knowledge with each other or with experts. Indeed, usually users and/or communities access a DL in search of some information. This means that it is quite possible that users may have overlapping interests if the information available in a DL matches their expectations, backgrounds, or motivations. Such users might well profit from each other's knowledge by sharing opinions or experiences or offering advice. Some users might enter into long-term relationships and eventually evolve into a community if only they were to become aware of each other. CYCLADES is indeed a DL environment supporting collaboration and personalization at various level, where users and communities may search, share and organize their information space according to their own personal view and where the system generates recommendation of various types based on user and community profiles.

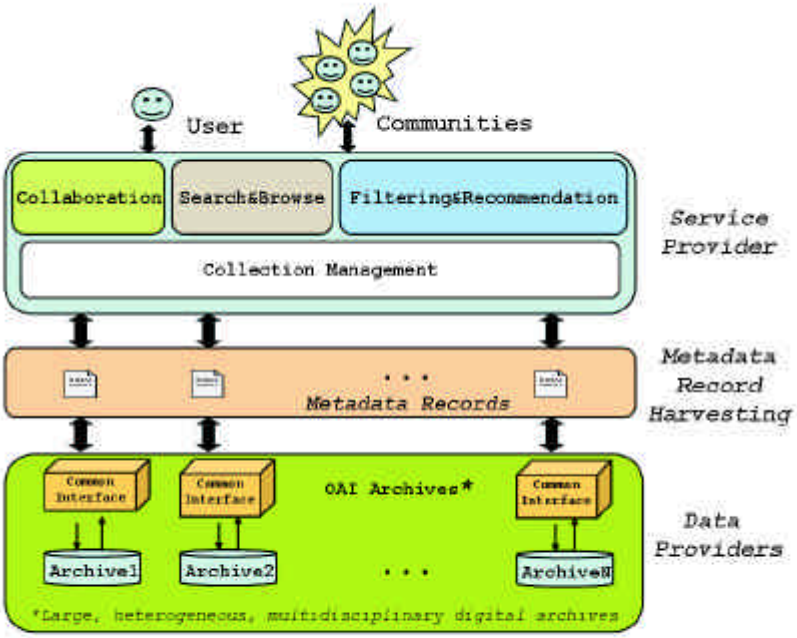

Figure 1: Logical view of CYCLADES.

A.2 A personalised collaborative DL environment

CYCLADES provides an integrated environment for users and groups of users (communities).

The logical view of its functionality is depicted in Figure 1. The digital archives to which CYCLADES users have access to are those adhering the *Open Archives Initiative* (OAI). Informally, the OAI is an agreement between several digital archives providers in order to provide some minimal level of interoperability between them. In particular, the OAI defines an easy-to-implement gathering protocol over HTTP, which give *data providers* (the individual archives) the possibility to make the documents' metadata in their archives externally available. Indeed, the agreement specifies that each document of an archive should posses a *metadata* record describing the documents properties and content. In particular, the format of the metadata records should be Dublin Core. The metadata record consists of several attributes describing author, title, abstract, etc. of documents. The protocol allows then to gather these metadata records (in place of the real documents). A link to the 'real' document is also present if the document is accessible. A metadata record may be understood as a statement of existence and short description of a document, which may be then accessible to a user according to the access policies of the archive, which owns the document.
To date, there is a wide range of archives available (more than one hundred registered archives) in terms of its content, forming a quite heterogeneous and multidisciplinary information space. The availability of the metadata records from the OAI compliant archives makes then it possible for *service providers* to build higher levels of functionality. In this sense, CYCLADES allows the access to the metadata provided by these archives, as it gathers these records, and through them provides access to the described documents (if they exist and their access is allowed). On top of them,

CYCLADES acts as an OAI service provider providing functionality for (i) advanced search in *large, heterogeneous, multidisciplinary digital archives*; (ii) collaboration; (iii) filtering; (iv) recommendation; and (v) the management of records grouped into *collections*. These functionality are available in several environments described below.

The *Collaborative Work Environment*, which is an extension of the BSCW environment (Basic Support for Collaborative Work), provides a folder-based environment (Figure 2 shows the content of a user folder, in our case the 'Physics-Gravity' folder of the community of physicists) for managing *e.g.* metadata records, queries, collections, external documents, ratings and annotations. In particular, users may organize their own information space according to their own hierarchy of folders.

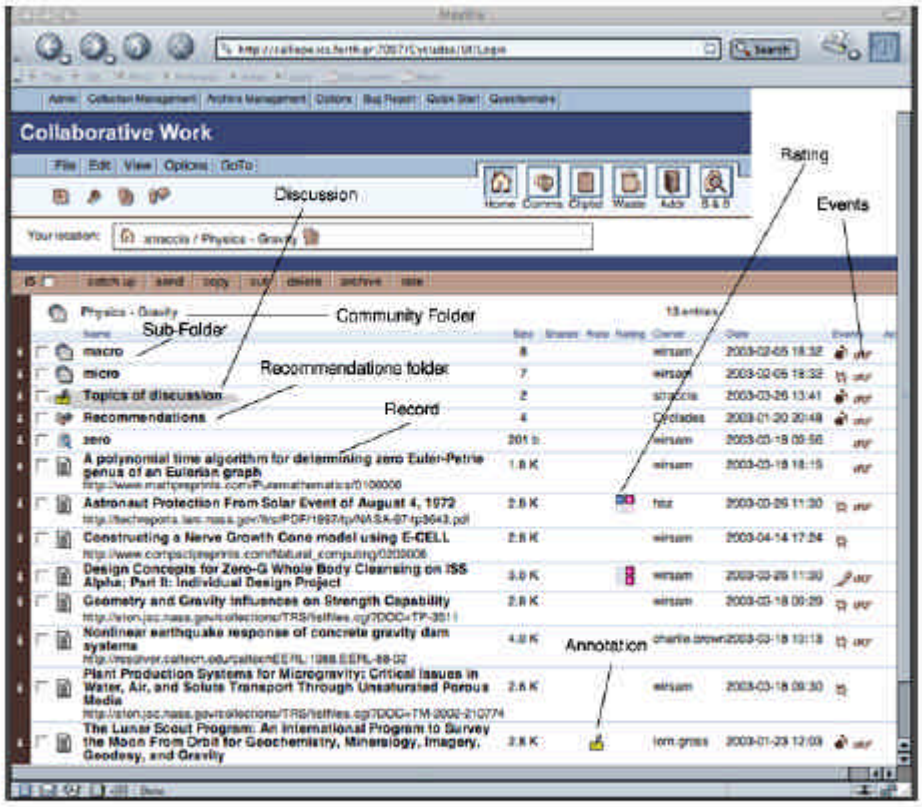

Figure 2: User interface: folder content

Each folder typically corresponds to one user related subject (or discipline, or field), so that it may be viewed as a thematic and usually semantically related repository of data items. There are two types of folders: (i) *private folders*, *i.e.* a folder owned by one user only. This kind of folder can only be accessed and manipulated by its owner. They are invisible to other users; and (ii) *community folders*, which can be accessed and manipulated by all members of the community that owns the folder. Community folders are used to share data items with other users and to build up a common folder hierarchy (rating, annotating, downloading and uploading of data items is permitted). Community folders may also contain *discussion forums* where notes may be exchanged in threaded discussions (similar to news groups). In order not to lose shared activity in the collaborative DL environment, mutual awareness can be supported through event icons displayed in the environment. Activity reports that are

daily received by email are also be possible. Users can also view the list of all existing communities and can join a community directly if this is allowed by the community's policy, or contact the community administrator in order to be invited to the community. In the collaborative work environment, the access policies can be set-up, as well as the notification (alerting) modalities.

The *Search and Browse Environment* supports the activity of searching records in the various collections accessible from within CYCLADES as well as to search into the shared folders or private folders a user owns. Users can issue a query and are al lowed to store selected records within their folders and community folders they have access to. Essentially, three types of search are supported: (i) in *ad-hoc search* a user specifies a query and the system looks for relevant records within a specified collection; (ii) *filtered search* is like the usual ad-hoc search, except that the user specifies, additionally to a query (*e.g.* 'zero'), also a target folder (*e.g.* 'Physics-Gravity'). The goal of the system consist then to find documents not only relevant to the query, but also relevant to the topic of the target folder (in our example, the request is something like 'find records about zero gravity'); and (iii) in *what's new, on-demand*, the user specifies a target folder, without specifying a query, and the goal of the system consists in finding all records, relevant to the target folder, which where become available to the system since the last time the user asked for this request. This corresponds roughly to the functionality provided by alerting services, except that the profile is build implicitly from the folder content, and that records are delivered to the user on-demand. The recommendation environment provides the off-line version.

The *Filtering and Recommendation Environment* supports the personalized search and provides the recommendations functionalities. All recommendations are specific to a given user folder (topic of interest), *i.e.* they have always to be understood in the context not of the general interests of the user, but of the specific interests (topic) of the user represented by a folder. A user may get recommendations of *metadata records* (suggesting to the user to access to relevant documents), *collections* (suggesting to the user to search within a relevant information space), *users* (suggesting to the user to enter in relationship with a user or give a look to the publicly available documents of the recommended user), and *communities* (suggesting to the user to join the community) issued to users based on user and/or community profiles. For instance, Figure 3 shows the recommendations related to the 'Physics-Gravity' folder, deemed by the system as relevant to this folder.

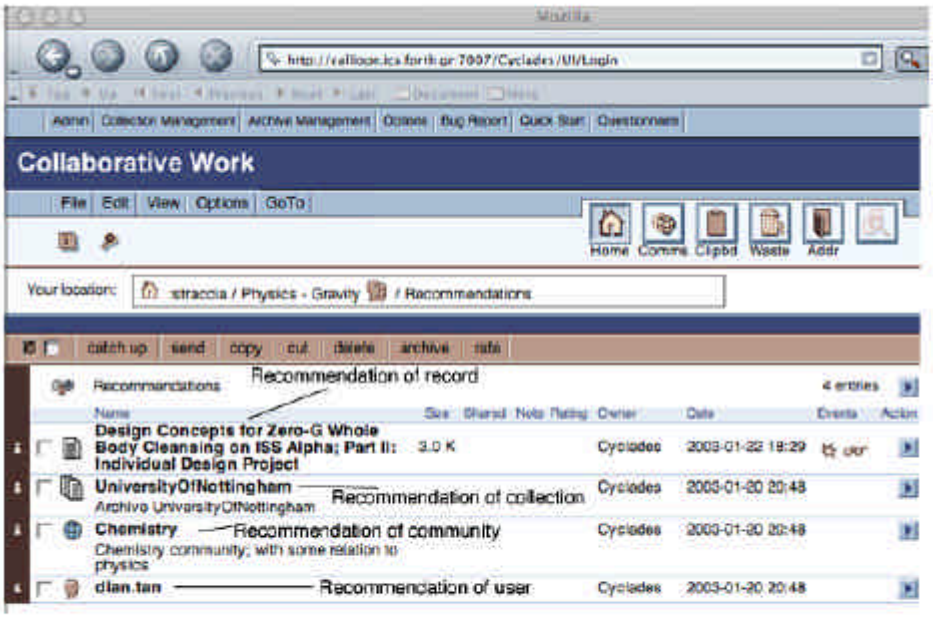

Figure 3: User interface: folder content and recommendations.

Finally, the *Collection Management* manages collections (*i.e.* their definition, creation, and update). Its aim is to allow a dynamic partitioning of the information space according to the users' interests, where to search into. Usually, a collection is meant to reflect a topic of interest of a user or a community, *e.g.* the collection of records about 'Information Retrieval'. Informally, a collection specification is the definition of a not materialized view over the information space and it is up to the system to automatically determine the archives in which to search for relevant records (this is accomplished by means of a technique called *automated source selection*.

Appendix B

This is the list of organisations that have been contacted in order to promote the course.

Library associations

AAB (Asociación andaluza de bibliotecarios) ANDALUSIA
http://www.aab.es

ABF (Association des bibliothécaires français) FRANCE
http://www.abf.asso.fr

BAD (Associação Portuguesa de Bibliotecários,Arquivistas e Documentalistas) PORTUGAL
http://www.sdum.uminho.pt/bad/

BAILER (British Association for Information and LibraryEducation and Research) UK
http://www.staff.livjm.ac.uk/busjofar/bailer/

BBS (Verband der Bibliotheken und derBibliothekarinnen/Bibliothekare der Schweiz = Associationdes bibliothèques et bibliothécaires suisses) SVIZERLAND
http://www.bbs.ch

BF (Bibliotekarforbundet = Union of Danish Librarians) DENMARK
http://www.bf.dk
Col·legi Oficial de Bibliotecaris-Documentalistesde Catalunya SPAIN-CATALOGNA
http://www.cobdc.org

VDB (Verein Deutscher Bibliothekare) GERMANY
http://www.vdb-online.org
IFLA (International Federation of Library Associations andInstitutions)
http://www.ifla.org

SBS (Svenska bibliotekariesamfundet) SWEDEN
http://www.bibliotekariesamfundet.se

VÖB (Vereinigung österreichischer Bibliothekarinnenund Bibliothekare) AUSTRIA
http://voeb.uibk.ac.at

VVBAD (Vlaamse Verening voor Bibliotheek-, Archief- enDocumentatiewezen) BELGIUM
http://www.vvbad.be

Archives and archive associations

ANAI (Associazione nazionale archivistica italiana)
http://www.anai.org

FIAF (Fédération Internationale des Archivesdu Film = La Federación Internacional de ArchivosFilmicos = The International Federation of Film Archives)
http://www.fiafnet.org/

International Television Association
http://www.itva.org/

On-line Digital libraries magazines

Diario Digital (http://www.diariodigital.pt)
Wissen Digital (http://v.hbi-stuttgart.de/Bibliothek/wd/)
Journal of Digital Information (http://jodi.ecs.soton.ac.uk/)
Transfroming Traditional Libraries (http://www.lib.usf.edu/mdibble/ttl/)
D-Lib Magazine (http://www.dlib.org/)
School Library Journal Online (http://www.slj.com/)
IFLA Journal (http://www.ifla.org/V/iflaj/index.htm)
Biblio Tech Review (http://www.biblio-tech.com)
Library Journal (http://libraryjournal.reviewsnews.com/)
Primary Source (http://www.imls.gov/whatsnew/new_imls.htm)

Other mailing lists

ECDL conference
OAF – Open Archive Forum
DELOS Network of Excellence

Appendix C

This appendix contains the flyer used to promote the course