

Second Progress Report

1 August 2001 – 31 January 2002

*Editor : B. Le Dantec – ERCIM
Remi Ronchaud
2004, route des Lucioles
B.P. 93
06902 Sophia Antipolis
France*

[Bruno.le_dantec@ercim.org](mailto: Bruno.le_dantec@ercim.org)
[Remi.ronchaud@ercim.org](mailto: Remi.ronchaud@ercim.org)

Progress Report 2

Covered period: From 1 August 2001 to 30 January 2002

Project Scientific Manager

Name: Umberto Straccia

Address: Istituto di Elaborazione dell'Informazione

Consiglio Nazionale delle Ricerche Via Moruzzi,1 , 56124, Pisa (Italy)

Phone: 39. 050 315 2894 Fax: 39. 050 315 2810

e-mail: straccia@iei.pi.cnr.it

Management report editor

Name: Bruno Le Dantec

Address : 2004 Route des Lucioles - BP 93 - 06902 Sophia Antipolis cedex - France

Phone: 33 4 92 38 50 13 Fax : 33 4 92 38 50 11

e-mail: bruno.le_dantec@ercim.org

1. EXECUTIVE SUMMARY

1.1 Project structure

Project organisation and synergy have been initiated through several meetings attended by all partners. All advances and results of the projects are recorded in the common project internet web pages, <http://www.ercim.org/cyclades>, which contain all reports, working documents, deliverables and other useful information. Such a system facilitates communication among partners, each of them being aware of the state of the project at any time. The project Technical Annex is also available from the web site for all the partner have an easy access to the project roadmap.

1.2 Scientific Progress

The main goal of the second six months of the project was to work on WP3, System Development. In particular the objective was both to start with the implementation of the various CYCLADES services, as well as to work on the definition of the various algorithms and database schema involved. Worth mentioning are the metadata record indexing algorithm in the Access Service, the filtering algorithms and the recommendation algorithms in the Filtering and Recommendation Service. The detailed specification of the Services will be part of Deliverable 3.0.1, which will be ready at the beginning of February. A draft version is currently available on the web site.

1.3 Results

The goals after the first twelve months of the project were:

- to collect and analyse end users requirements;
- to define and prepare the functional and architectural specifications of Cyclades
- to ensure close cooperation with the Open Archive Initiative
- to disseminate information about the project;
- to manage both scientifically and administratively the project.

These objectives were achieved as all the corresponding deliverable were produced.

- Progress Report 1 (D9.2.1)
- Dissemination and Use Plan (D7.2.1)
- Coordination Report (D8.1.1)
- Global Architecture report (D2.2.1)
- Quality Insurance Plan (D9.1.1)
- Project Web site report (D7.1.1)
- User Requirements Report (D1.2.1)

A lot of effort was spent in carefully specifying the system architecture (D2.2.1) in order to provide the implementation and integration work with reliable information.

2. MANAGEMENT

2.1 Meetings

Regular meetings, sometimes informal, were held to maintain a general cohesion within the project. Official Plenary Meetings are also organised when necessary to address very specific issues. The one held in the Fraunhofer FIT institute in Bonn on the 29 and 30 of November 2001 was organised to address:

1. the current implementation state
2. discuss the organisation of review in January
3. the preparation of Deliverable D3.0.1 due to early February.

Concerning the above points, especially for case 3 above, CNR being the WP3 leader, it has suggested the preparation a preliminary table of contents, just to have a starting point.

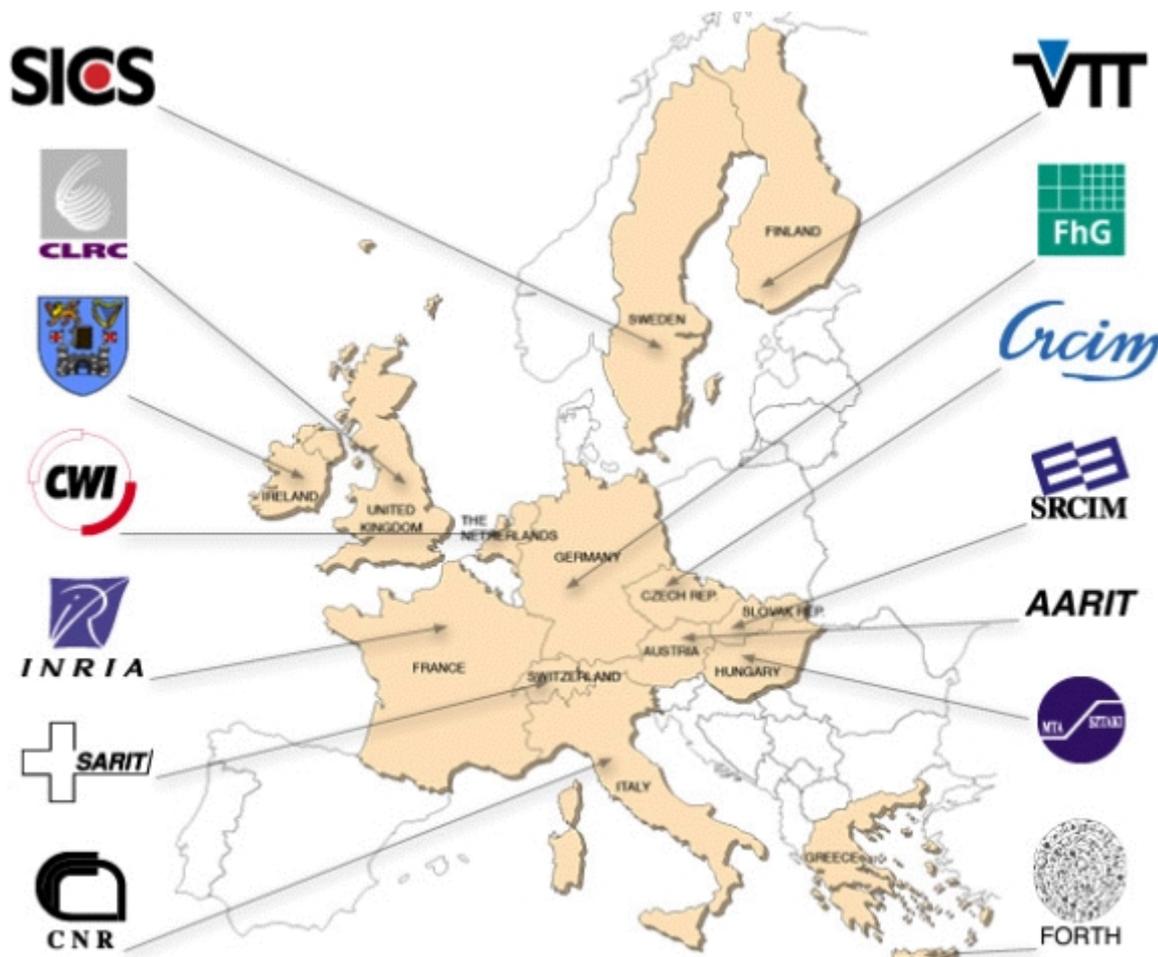
2.2 Web Server

The web site was regularly updated with the final versions of the Deliverables and Reports. In addition, the entire site has been redesigned as it was generally agreed that it was too dim and did not look attractive enough. Beyond the packaging facelift, new services are being provided to the visitors, such as a specific button to reach the project or a small search engine to find the documents or partners.

2.3 Presentations in Conferences and Exhibitions and Public Relations

The Cyclades fact sheet organised as a leaflet are regularly distributed to various meetings and Conferences to enhance Cyclades visibility. For example, during the IST Conference held in Dusseldorf on the 3,4 and 5 December 2001, ERCIM has distributed several copies of this fact sheet, not only to representatives of the European Commission but also to NSF delegates and several heads of research Institutes in Europe.

In addition, ERCIM has distributed a pack of these fact sheets to all its members institutes in Europe for dissemination:



Finally, CNR made a Cyclades presentation at the Italian National Workshop on "Digital Libraries: Research Development in Italy, on the 7 - 12 October 2001.

2.4 Publications

Tom Gross from Fraunhofer FIT produced "*Sharing and Exchanging Knowledge in Communities: The CYCLADES Open Collaborative Virtual Archive Environment.*" This paper was presented at Workshop on Community Knowledge at the Seventh European Conference on Computer-Supported Cooperative Work - ECSCW 2001 (Sept. 16, Bonn, Germany). Nakata, K. and Kuutti, K., eds. 2001. pp. 17-24.

The ERCIM News magazines distributed over 7000 copies every 2 months provides a powerful media support. When relevant, a specific article is published in the two pages reserved for Digital Libraries activities.

3. ACTIVITIES AND ACHIEVEMENTS

3.1 ERCIM

During last six months of this first year ERCIM has been actively involved in WP 7 and 9, respectively Dissemination and Management.

Activities & Achievements

First of all, ERCIM main focus was still set on providing the consortium with the most appropriate support through dedicated tools or through a coherent project management policy.

Therefore in WP9, ERCIM, along with CNR, has been paying particular attention to improving the self-assessment of the Project. The various documents due to the Commission are presented to all partners on the BSCW site for reviewing. Any document can be read, opened or modified. The final version of the documents and reports are then available on the ANFAS web site, under the document repository section.

Regarding the coordination of the work performed by the Cyclades consortium, ERCIM has brought an additional degree of quality in the work achieved through the evaluation and review of all the strategic documents. Some of these documents being necessary for some partners to move onwards in the project, it was the coordinator's priority to ensure that such documents fulfilled their role, especially internal roadmap documents that are the cornerstones of the project.

In that perspective, the quality of each document produced was reviewed not only by the workpackage leaders and the Scientific Co-ordinator, but also by the person using this document as an input. If it was not regarded as satisfactory, the document was sent back to its author, precisating what elements needed further details. Additional proof-reading was finally done to give the reports its ultimate polish.

This system, operational from the beginning of the project, was applied to all documents, especially to Deliverable that needed to be reviewed and assessed by all key players of the project before being sent to the European Commission.

The collection of every partners' contribution to the Progress Reports and Man-month effort tables was another co-ordination task carried out. These tables have also been revised and modified in accordance with European Commission's recommendations (improved tables and graphics) for a better readability. The Management integrated these elements to realise the documents to be sent to the Commission, only after a general partners' assessment of the reports. The Management was also gathering the various detailed Cost Summary to hand in the Cost Statement consolidated figures.

Additional efforts were made by the Management to strengthen the links with other digital Library oriented projects (DELOS (IST-1999-12262), SCHOLNET (IST-1999-20664),...) with the aim to foster pragmatic project clustering

In addition, the Technical Coordination followed closely the achievements of each WP individually and brought every partner to further interact with the different WP leaders accordingly to the work programme. Beyond this issue, since all Deliverables and Reports achieved were reviewed by the partners, they were also required to provide information and feedback to the project management. The outcome here being to define more precisely the general course of action for Cyclades.

Regular technical meetings were also organised to ensure closer collaborative work among the different Cyclades teams involved in the project. In this perspective also, a preparatory meeting will be organised in Luxembourg on the 20th of January 2002. This will be a rehearsal to the project's first review and the opportunity for all partners to co-ordinate their actions for new rest of the project.

Regarding the Dissemination activities and achievements of WP7, particular efforts have been made to keep the Cyclades web site regularly updated. While the BSCW co-operative work platform hosted the deliverables and reports "in progress", the Cyclades web site document repository with restricted access hosted the final and validated version of the documents, similar to those sent to the European Commission.

As for the visibility of the project itself, the web site was totally redesigned to a more professional and modern style. The logo was not abandoned in order for the project to keep its "identification mark" but the fact sheet was redesigned also to match the web site general layout.

Regular information about the project was also given through a specific Digital Library Double page in the ERCIM News magazine , more than 7000 copies of which are distributed every 2 months.

Finally, all the Cyclades partners have been oriented towards increasing their participation to Conferences, workshops and to increase Cyclades related publications to foster the project visibility world-wide, bearing in mind to reach also the potential end-users: archives, libraries,

3.2 CNR

During the last six months the CNR has been proceeding with the work involved in developing the Collection Service and Filtering and Recommendation Service for the CYCLADES system prototype (WP3). CNR has also proceeded with coordination of the activities between CYCLADES and OAI (WP8), as well as the overall scientific management of the project. These activities are described below.

WP3

CNR is responsible WP3 as a whole. Additionally CNR is responsible for the Collection Service (Task 3.3) and Filtering and Recommendation Service (Task 3.4, 3.5) of the CYCLADES system prototype. This phase concentrated on development of the detailed software specifications.

Actions taken during this reporting period:

Collection Service

- Development environment and base technology installed:
 - Apache XML-RPC (<http://xml.apache.org/xmlrpc>), MySQL database, Java
- Concept for database scheme developed
- Some methods implemented
 - (addCollection, listCollection)

Filtering and Recommendation Service

- Development environment and base technology installed (XML-RPC API, Java, BerkleyDB)
- Concept for database scheme developed
- Filtering algorithms defined
- Filtering algorithms implementation in progress
- Recommendation algorithms defined

CNR also contributes significantly to the definition of the Detailed System Specification (Deliverable D3.0.1).

WP8

The goals of WP8 are to Monitor progress achieved by CYCLADES and OAI. This monitoring should guarantee that:

- the interoperability agreements defined by OAI and their evolution is correctly understood and used by CYCLADES,
- the impact on the service level of new OAI specs is thoroughly evaluated and feedback is sent to the originators of the changes.
- Feedback to the OAI. This entails that validation and experimentation results will be made available to OAI for undertaking the necessary corrective actions

Meetings with OAI representatives. Regular meetings will be organised throughout the rest of the project to ensure a continuous monitoring.

Actions taken during this reporting period:

Monitoring progress achieved by CYCLADES and OAI.

- Continuous monitoring activity of the OAI (meetings and related papers) is in progress.
- The list of registered repositories as well as the register OAI services is continuously monitored.

Feedback to the OAI.

A first metadata gathering activity has been started:

- Some statistics about the to be gathered data from the OAI compliant archives has been collected.
- Some non-uniformity of the data has been detected.
- The report of this gathering process has been submitted to a OAI Workshop.
- The gathered data seems already to be enough for making the system appealing.

Meetings with OAI representatives

- CYCLADES will collaborate with the OAF (European Open Archive Forum) project, where IEI-CNR is a participant.

WP 9

The CNR team has been ensuring the scientific management of the entire project. During the last six months CNR monitored particularly the timing and scientific progress of WP3 (and promoted a technical meeting in Bonn), as well as the preparation of the first CYCLADES review meeting.

3.3 FORTH

For the reporting period, August 2001 to January 2002, ICS-FORTH undertook the task of providing a detailed specification of the Mediator Service of the CYCLADES system prototype. The Mediator Service is the service of the CYCLADES system that integrates all participating services and enables the communication among them. It also supports the global system user interface and controls the registration of users and new services to the system.

More specifically, ICS-FORTH specified and produced a first implementation of the classes and methods that constitute the Mediator Service and which support the addition, removal or update of services, service discovery primitives, error logging and reporting, addition of users, user information logging, access rights management and inter-service communication primitives. A database storing service and user information has been designed and implemented in the Oracle8i relational database management system. The database maintains information about the registered

users of the systems and their profiles. Communication of the Mediator Service and the database is implemented through JDBC.

The design and implementation of the Mediator service treats all system components homogeneously. The current implementation renders the system open and extensible, i.e., more services or service instances can be added at any time. The implementation has been carried out in Java using servlet technology and is portable to different platforms.

ICS-FORTH also participated in the detailed specification of the communication protocol for the system services. Specifically, the encoding of the service input/output parameters into the chosen communication protocol (XML-RPC) has been carried out in collaboration with Fraunhofer FIT. This task also included the mapping of XML-RPC data types to the respective data types of the Java language as well as, the specification of the unified fault codes to be used in the entire system. A comparison of XML-RPC to SOAP – an XML-based generic communication protocol – was carried out in order to validate the choice of XML-RPC as the protocol to be used within CYCLADES.

Dimitris Plexousakis and Nick Papadopoulos of ICS-FORTH participated in the meeting that took place in Bonn in December 2001.

Dimitris Plexousakis attended the VLDB 2001 conference in Rome, Italy in September 2001. Services over distributed environments such as the one of CYCLADES are an emerging topic in this major forum of database research. The works reported in this forum are useful in assessing the design and architectural choices made so far in the progress of the project and in understanding the state-of-the art in distributed service environments,

3.3 Fraunhofer FIT

Fraunhofer FIT is the leader of Workpackage 4, starting at t+19. As this has not started yet, most of the effort of Fraunhofer FIT has been put into Workpackage 3.

Detailed system specifications were written for the Collaborative Work Service (CWS) and the Rating Management Service (RMS). They contain detailed information on the functionality, the process flow, the internal architecture, the data and methods involved, the user interface, and the interaction with other services for both services.

In preparation of the implementation of the CWS and the RMS the base technologies have been installed and tested: a XML-RPC server using F. Lundhs xmlrpc-lib from Secret Labs AB; a BSCW server from OrbiTeam; an Apache Web server; a MySQL database; etc. ...

Concerning the CWS the classes for private, project, and community folders, for records and queries, for users, communities, and collections were modelled and partly implemented. The user interface was designed and partly implemented. Concerning the RMS the database scheme for the ratings was developed and implemented. Some methods of the RMS have been made available for the other CYCLADES partners, so they can try out parts of the functionality in real-time (cf. http://CYCLADES.gmd.de/cgi-bin/cyc_cws.cyc).

Fraunhofer FIT has organised a meeting of the CYCLADES partners in St. Augustin on 29-30 November, where the communication protocol, the APIs of the individual services, and the parameter encoding and error codes were discussed. Fraunhofer FIT has then written the precise specification of the communication protocol in XML-RPC and the parameter encoding including data types, object IDs, and the mapping of XML-RPC data types to the programming languages mainly used in CYCLADES (Java and Python).

Fraunhofer FIT has participated in a workshop on Community Knowledge at the Seventh European Conference on Computer-Supported Cooperative Work - ECSCW 2001 in Bonn.

3.5 UNIDO

Activities:

During the whole period covered by this report, UniDo has been working on Dortmund's contributions to WP3.

As the Cyclades system will harvest metadata from electronic sources that adhere to the Open Archives specification, and will furthermore provide value-added services on top of this metadata, the Cyclades system can be seen as an Open Archives service provider. UniDo, as the Cyclades partner who is developing the harvesting component, is interested in keeping close contact with the Open Archives community.

Achievements:

For the services to be developed at Dortmund, a database system and programming tools were chosen, installed, and tested. UniDo will use Perl, Java, and Tomcat for the prototype development. The database system chosen for the Access Service prototype is Postgres, but the internal architecture of the service has been designed in such a way that Postgres be easily substituted by other relational database systems.

An abstract syntax both for the formulation of queries, and for the definition of collection filtering conditions was defined, allowing mandatory as well as optional, weighted query conditions.

The internal architecture of the Access Service and the Search and Browse Service was specified. An important design consideration was modularity, so that individual components can be easily adapted or substituted later.

The first component of the Access Service, the harvesting and indexing component, has been implemented. UniDo is currently running the component in a test environment.

One problem that UniDo came upon during the first test harvestings is the internal heterogeneity of the metadata provided by the different archives. The Open Archives protocol specifies a uniform way of accessing the data, and one common format for metadata records as minimum, but it does not further define the content of the fields inside the metadata records. This leads to very heterogeneous data which results in difficult indexing and retrieval. UniDo has made several suggestions to alleviate this problem and presented them in a short paper at the ECDL 2001 Workshop On Experimental OAI Based Digital Library Systems in September 2001 in Darmstadt.

4. OTHER

5. DEVIATION FROM PLAN