

# An open collaborative virtual archive environment

Umberto Straccia, Costantino Thanos

I.S.T.I - C.N.R., Area della Ricerca, Via G. Moruzzi, 1, 56124 Pisa, Italy  
e-mail: {umberto.straccia, costantino.thanos}@isti.cnr.it

Published online: 23 July 2004 – © Springer-Verlag 2004

**Abstract.** CYCLADES is a system that combines several technologies from the information retrieval and digital library fields, where users and user communities may deal with a rather large set of heterogeneous digital archives. CYCLADES provides a highly personalized environment where users may organize (and search) information space according to individual taste and use; in addition, CYCLADES provides advanced features of collaborative work among users.

**Keywords:** Personalization – Collaboration

It is widely recognized that the Internet is growing rapidly in terms of the number of users accessing it and the amount of *digital libraries* (DLs) created and accessible through it. This has made it increasingly difficult for individuals to control and effectively seek information from the potentially infinite number of DLs available on the Internet. A common characteristic of most DLs is that *they do not provide personalized support to individual users*. In practice, users use the same information resources over and over again and would benefit from advanced search and browse features to reduce the repetition of manual labor expended only to find documents that are identical to the first ones. Additionally, users are highly interested in being able to organize information space according to their own subjective perspective. The requirement of personalized search in the context of DLs is already known, and some DLs provide related functionality ([1–5]). Many of them fall into the category of *alerting services*, i.e., services that notify a user (by sending an e-mail) with a list of references to new documents deemed to be relevant. Typically, all these services are based on the so-called notion of *user profile* (a machine representation of the user's information need), acquired either automatically (by user–system interaction) or set up manually (by the user) and used to *recommend* relevant data.

Very rarely are DLs considered as *collaborative meeting* or *working places*, where users may become aware of each other, open communication channels, and exchange information and knowledge with each other or with experts. Users usually access a DL in search of some information. This means that it is quite probable 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 by 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. Such a service could be important for a DL as it supplies very focused information.

CYCLADES<sup>1</sup> is indeed a DL environment supporting collaboration and personalization at various levels, where users and communities may search, share, and organize their information space according to their own personal view and where the system generates recommendations for various types based on user and community profiles. The logical view of its functionality is depicted in Fig. 2. The digital archives to which CYCLADES users have access are those adhering to the *Open Archives Initiative* (OAI).<sup>2</sup> Informally, the OAI is an agreement between several digital archive providers to provide some minimal level of interoperability among them. The availability of the metadata records from the OAI-compliant archives then makes it possible for *service providers* to build higher levels of functionality. In this sense, CYCLADES grants access to the metadata provided by these archives and through them access to the represented document (if it exists and the access is allowed). Informally, the various environments are the following. The *collaborative work environment* provides a folder-based

<sup>1</sup> [www.ercim.org/cyclades](http://www.ercim.org/cyclades)

<sup>2</sup> [www.openarchives.org](http://www.openarchives.org)

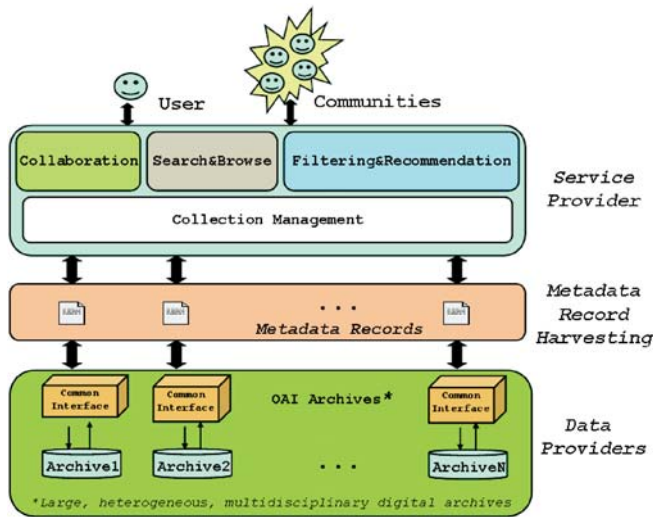


Fig. 1. Logical view of CYCLADES

environment (Fig. 2 shows the content of a user folder) for managing metadata records, queries, collections, external documents, ratings, and annotations. In particular, users may organize their own information space according to their own folder hierarchy, i.e., a set of hierarchically organized folders, each of which is a repository of a user's selected data items. Each folder typically corresponds to one subject (or discipline or field), so that it may be viewed as a thematic repository of data items (users may save data items to folders). 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 hier-

archy (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 e-mail are also possible. Users can also view the list of all existing communities and 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.

The *search and browse environment* supports the activity of searching records in the various collections, formulating and reusing queries, and browsing schemas, attribute values, and metadata records. Users can issue a query and store selected records within their folder hierarchy. In an *ad hoc search* a user specifies a query and the system looks for relevant records within a specified collection. We further allow a kind of *filtered search*: this is like the usual ad hoc search, except that the user specifies a query and a folder and the system looks for records relevant to both the query and the folder.

The *filtering and recommendation environment* allows for a filtered search and provides recommendations. A user may get recommendations of *data items*, *collections*, *users*, and *communities* issued to users based on user and/or community profiles. All recommendations are specific to a given user folder (i.e., topic of interest), that is, they always have to be understood in the context not of a user's general interests, but of the specific interests (topic) of the user represented by a folder.

The *collection management* manages collections (i.e., their definition, creation, and update), thus allowing a partitioning of the information space according to the user interests.

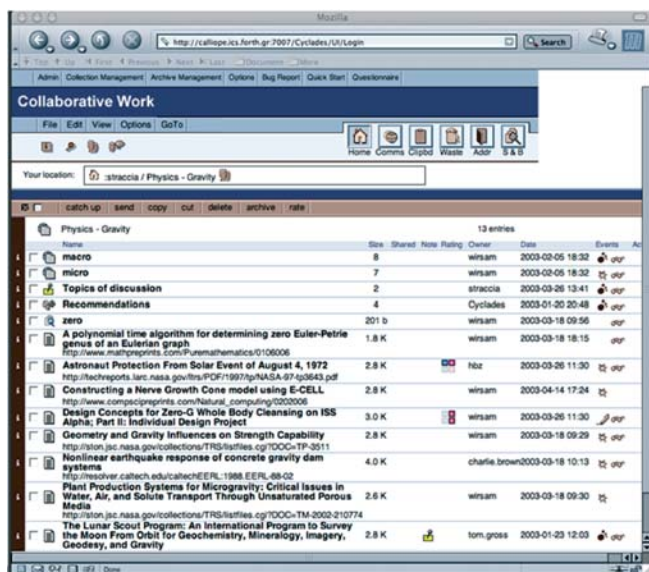


Fig. 2. User interface: folder content

## References

1. Bollacker K, Lawrence S, Giles CL (1999) A system for automatic personalized tracking of scientific literature on the web. In: Proceedings of the ACM conference on digital libraries, Berkeley, CA, 11–14 August 1999. ACM Press, New York, pp 105–113
2. Di Giacomo M, Mahoney D, Bollen J, Monroy-Hernandez A, Rouiz Meraz CM (2001) Mylibrary, a personalization service for digital library environments. In: Proceedings of the Second DELOS Network of Excellence Workshop on Personalisation and Recommender Systems in Digital Libraries, Dublin City University, Ireland, 18–20 June 2001
3. Faensen D, Faulstich L, Schweppe H, Hinze A, Steidinger A (2001) Hermes: a notification service for digital libraries. In: Proceedings of the ACM/IEEE joint conference on digital libraries, Roanoke, VA, 24–28 June 2001, pp 373–380
4. Fernandez L, Sanchez JA, Garcia A (2000) Mibiblio: personal spaces in a digital library universe. In: Proceedings of the ACM conference on digital libraries 2000, San Antonio, TX, 2–7 June 2000, pp 232–233
5. Rocha LM (1999) Talkmine and the adaptive recommendation project. In: Proceedings of the ACM conference on digital libraries, Berkeley, CA, 11–14 August 1999, pp 242–243