TORII
# Access to the Digital Research Community

Marco Fabbrichesi

INFN/SISSA
via Beirut 4
34014 Trieste
Italy

The communication of the results of scientific research and in many ways research itself have changed in recent years as digital means of information production, distribution and access have become widespread.

Think of your desk, or your office. The way you work has already changed. Paper preprints have been replaced by electronic archives,
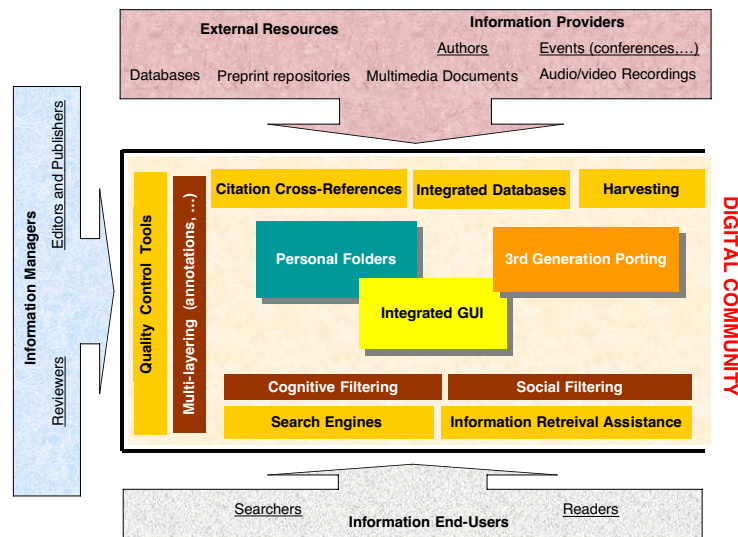
mail and phone calls by e-mail, typewriters and hand drawing by text and graphics software programs, cabinet files by saved directories on hard disks. These new tools, together with multimedia presentations and conference websites, constitute the growing digital network of information that is taking over many aspects of the working place of research. It is a system in which the information flow is regulated, integrated and made available by the software and the network.

The digital network of research is currently organized in three layers:

I Repositories of information: open archives and databases. This first level is the analogous of library and publishers stacks.

II Services over and for information: e.g. review journal, cross-citation. They are the analogous of, for instance, library desks and paper journals.

III Digital communities: synergic union of services and information. Ideally, they replace your desktop environment by giving access to the tools you use in your everyday work.
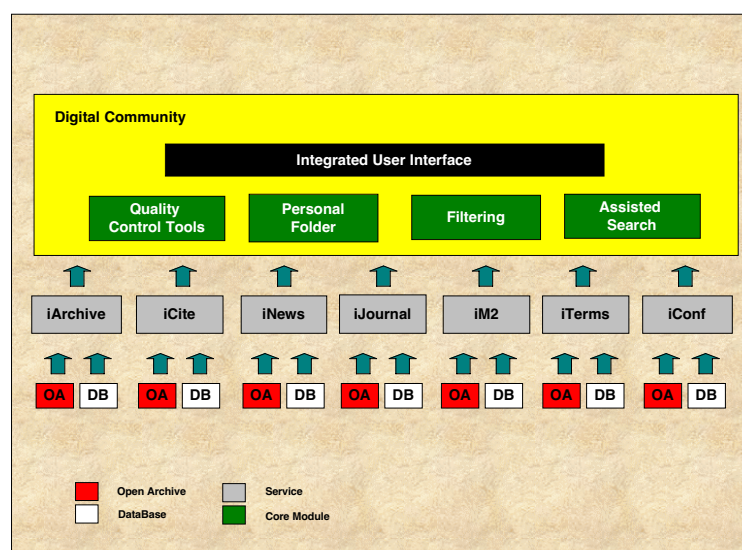
Only the first two have been fully digitalized: Torii is the first attempt to complete the structure.

Torii at `torii.info` gives you direct access to the digital research community. It works the way you work. All tools and documents you need are collected under an unified access point, organized according to your needs and ready for you everywhere you are and at any time you may need them. An intuitive user interface helps you to navigate. All the tools you need are at your finger tips. Choice of archives and subjects are easily costumized to fit your interests.

And the platform grows as the digital community grows. New features will be added as they become available in the future.

The personal folder is the hub of the system. You can store your documents here for future reference or to be printed or sent to others. The personal folder is easy to use by means of its drag-and-drop interface. It ideally replaces the cabinet filer where paper documents used to be stored. Stored documents can be ranked according to your profiles, impact factors or evaluation tools.

And there is more: you will find in your personal folder new documents suggested by the social filtering engine and you can attach to any documents comments for yourself or to be shared by the community.

The multi-layered document is a stack of documents that we want to manipulate. It could be an entry in a database, and as a new layer is added so is the entry column modified in the database, or it could be a collection of documents managed by a web server that keeps track of their relationships and modifications. The access to a multi-layered document is dynamical. According to who you are at a given moment—reader, author, referee, editor—you have access to different layers. Dynamical access requires an appropriate interface between the multi-layered documents and the users. It also requires intelligent agents to sift through the increasingly large amount of information to shape it into some hierarchy and thus making it usable.

Key features for the integration of dynamic access to the information into the portal are the XML language and the Open Archive Initiative protocol. The XML language is used to encapsulate in a common structure the exchanged information, originally stored in a variety of formats. This XML metadata structure will represent the semantic aspects related to the data.

The Open Archives Initiative Protocol for Metadata Harvesting defines a HTTP-based mechanism for harvesting XML files containing metadata from repositories.

This is the basic communications protocol between Torii and the underlying services, operating in a location-transparent way. The Open Archives Initiative, indeed, develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content. The goal of the Open Archives Initiative is to supply and promote an application-independent interoperability framework that can be used by a variety of communities engaged in publishing content on the Web. Through the use of the Open Archive Initiative protocol, Torii will be easily extensible to any archive implementing the protocol.

In a user-friendly information society, the information overload is limited and the information delivery is personalized: the broad-casting of information is replaced by a more effective narrow-casting and mass-media are replaced by personal media tailored to each user's needs. These aims can be reached by more effective systems for information access. Torii provides a filtering component to skim too large a set of retrieved information and thus providing the user only with the information nearest to his interests.

The user defines his research interest profiles by filling in a form; from this a user profile is derived, based on a semantic network. The profile is automatically updated every time the user provides explicit relevance feedback on some new documents. Documents to be evaluated by the cognitive filtering module are processed through information extraction techniques aimed at capturing the meaning of the document content. These techniques exploit linguistic processing and statistical analysis. Every day the filtering module filters the submissions to the archives accordingly to the user profiles and the graphical user interface provides tools to rank displayed documents accordingly to the user's profiles.

Out of the 30-50 daily submissions, the user is able to see the 3-4 most relevant at the top of the list.

Social filtering circulates interesting documents among users who share interests. It automatically feeds in the personal folder those documents that are potentially relevant for the user. The relevance of the documents is evaluated for similarity with the selections done by other users with similar interests. The process is the digital analogous of sharing paper among collegues. It fosters the growth of the digital community.

Quality control tools memorize and exploit human evaluation of documents. They provide users with the possibility to express their evaluation of a document by filling in a predefined form and writing free textual comments. The form results are used to statistically evaluate numerical scores about the scientific quality of the document, the comments are general, each user can choose whether his comment will be public or for himself. Users can read all public comments on a document. These tools embody a first instance of open peer review in which the community as a whole participate in the review process.

A search engine, Okapi, is accessed directly from Torii. It offers a sophisticated search environment where you can look and search among the more than 150,000 documents currently stored in the archives. Okapi offers advanced retrieval mechanisms based on the probabilistic model of retrieval and relevance feedback. It runs on both the document metadata and their full text. It is fast and accurate.

You are not left alone in searching. An assistant monitors your search and helps you with helpful hints and terminological and contextual suggestions. It alerts you for dead-end searches leading to hundreds of documents or no document at all. You are made aware of strategic aspects of searching that allow you to fully exploit all information resources and services. The assistant comes fully integrated into the Okapi search engine of Torii.

Every day iCite extracts all citations from all the documents submitted to the archives. These are used to rank documents in Torii so that you can order them according to their impact factors. It is a completely automatic system that creates a net of cross-references inside the archives. It is an instance of service, the second level of the three-layer structure, that can also be accessed independently at `icite.sissa.it` to search for citations patterns and ranking.

The role of libraries in the near future will be that of maintaining archives and databases and providing access for its users to the digital communities by subscribing to them. Aside from collections of old paper resources, new documents accessed though archive and services providers will be available on screen and, on request, printed locally by the new generation of digital printing/binding machines. the budget nowadays spent on journal subscriptions will progressively be transferred to community subscriptions, archive mainteinance and hardware for users.

Torii is ready to move on into the future of digital networking. As the next generation of wireless systems comes into production, Torii will be accessible from your mobile phone. You can connect already and use it via WAP at

`torii.sissa.it/wml/ia.wml` but the full potentiality of the system must wait for the 3G broad bandwidth to come into being. At that point, you will be able to browse documents use your personal folder and any other of the features of Torii as you travel.