A Digital Library for an Academic and Research Community

by José Luis Borbinha & José Delgado *

Abstract
In the first part of this paper we discuss and define the concept of a networked digital library. We define it both as a new tool for virtual communities engaged in the production and dissemination of information and knowledge, and also as a potential active member of those communities.

In the second part of the paper we present the ARQUITEC project, a trial conceived to assess the defined concept of a networked digital library. ARQUITEC is a work in progress that will result in a prototype of a networked digital library for the Portuguese academic and research community.

Introduction
The actual and future impact of the Internet in our society is one of the most complex and participated discussions of the moment. An emerging issue of this discussion has been the redefinition of the role of libraries, raising the question of what is a "digital library" in a global networked world.

In the first part of the paper we discuss and define the Internet as a communication medium and as a meeting-place, a "new land" of opportunities for the virtual communities. This vision is discussed in opposition to a common vision of the Internet as just a new distribution medium, in the line of the press, the radio or the television.

Based on that discussion, we define the concept of a networked digital library. A networked digital library is seen not only as a repository for data and information, with the traditional missions of preservation and dissemination of knowledge, but also as an active partner with the potential to stimulate, support and register the process of creation of that knowledge.

In the second part of the paper we describe ARQUITEC, a prototype of a networked digital library for the Portuguese academic and research community.

ARQUITEC is a joint effort undertaken by INESC (an R&D institute), the Portuguese National Library and JNICT (the Portuguese R&D funding agency). The purpose of ARQUITEC is to set up a prototype of a networked digital library for the Portuguese research and academic community, which will be used to test the concept and the technology.

The vision
The net isn't 30 million people, it's tens of thousands of overlapping groups ranging from a few people to perhaps a couple of hundred thousand at the largest" (O'Rally, 1996).
It has been broadly pointed out that the information technology in general, and the Internet in particular, has been supporting the existence of virtual communities, defined as communities of individuals sharing common interests, but that are not geographically confined. Evident demonstrations of that reality are the existing thousands of News groups and electronic mail lists, dedicated to almost all the cultural, professional and political perspectives.

With that reality, the Internet can be defined as a new virtual space, like a new dimension of the physical and temporal world. It offers a real meeting-place and a multidimensional communication medium, with a social function in the genealogical line of the traditional squares, market places, coffeehouses (see the success of the cybercafes) and the telephone. This is a deeper and vaster view than merely defining it as a simple one-way broadcasting medium, such as the press, the radio or the TV, since in the Internet each one can be an equal player, with the same chances to be active as anyone else.

This vision has been already a field of concrete experiences in scientific and academic communities. It was maybe first identified by Paul Ginsparg and Steven Harnad, that coined expressions like "skywriting", "esoteric publishing" and "pre-print continuum" (Okerson, 1995; Harnad, 1990; Harnad, 1991; Harnad, 1995). Harnad presents an interesting perspective on the evolution of the human communication, with the phases of speech, writing, printing and, now with the Internet, skywriting. Skywriting is defined as both a new medium and a new model of communication, interactive, independent of the space and more suitable with the human cognitive process. This is a scenario favorable to the raising of esoteric virtual communities that, by using the Internet for their natural skywriting and pre-print activities, will be able to work and prosper in the production of their knowledge and memory.

With this reflection we can now complete the view of the Internet as the mean (the "ether") that can allow the library, now converted in the networked digital library, to go and meet the community. Networked digital libraries can be important not only for the geographically defined communities (that have already their traditional communal structures), but even more important for the geographically unbounded communities, where they can play as active members in the process of development and creation of knowledge and memory. Table 1 resumes that vision for the networked digital library paradigm.

In what we call the traditional library, the subject is the book. Its value is "sacred" (otherwise it wouldn't have been purchased) and it is stored "for ever". In this scenario authors decide what to write and when to edit the book, while the librarian decides whether to buy it or not. Finally, the librarians expect the patrons to came to the library and request the book. It was more or less like that until the middle of this century, when the industrial development changed it.

The industrial development reduced printing costs, illiteracy and the physical distances, while at the same time it increased the amount of information produced. It is not possible anymore for an individual to absorb all the knowledge produced by mankind, so it is necessary to specialize. The specialization brought thematic magazines, journals, reports, conference, etc. A new subject emergent from this reality is the "paper", which represents a new type of knowledge. It is not "sacred" anymore, but still formal, being validated by the credibility of an editor or a review committee. This knowledge is not intended to be valid "for ever", but to be discussed during a period of time, refined and, in the end, what survives is then sanctified in books (while the journals and conference proceedings are stored in the basement).

It is difficult for the traditional library to follow the specialization; so the library itself becomes specialized, with the mission to serve specific communities. Usually, those communities control now the library content in their own interest, in the sense of who decides which periodicals to subscribe or what to buy. Quoting Nicholas Negroponte:

> The real value of a network is more related with community than with information. The information super-highway is more than a shortcut to all the books in the Library of the Congress. It is creating a completely new global social tissue" (Negroponte, 1996).

In this scenario the library is requested to perform now a more active role. Since the communities are well identified, it is now possible to anticipate their needs and to provide customized services, such as the notification of new issues, advertisement of new publications, etc.

The scenario changes again with the arriving of the computer. With the desktop publishing tools and WWW, everyone becomes a potential publisher. The process acquires speed, and the subject is the idea. With computer networks, electronic mail and News groups, communities intensify their interactions. To produce fast results, ideas are submitted in pre-prints or presented to discussion as position papers in informal workshops. Ideas that succeed in this process are then published in journals and promoted in formal conferences. What will be the impact of this new reality in the library world?

Using electronic mail and WWW, it is easier for the library to reach the communities and provide new services (such as the announcement of workshops, the arriving of new publications, etc.). By the same reason, it is now easy for the users to interact with the library, not only to access Online Public Access Catalog (OPAC) services but, in an extreme scenario, to contribute also with new kinds of meta-knowledge that can enrich notably the library. Examples of such contributions can be the tuning and completing of thesaurus and catalogue (allowing dynamic and collaborative cataloguing), the attachment of annotations and comments to the stored documents (allowing collaborative refereeing, for example), etc.

Paradigms	**Networked Digital Library**		
	Specialized Library		
	Traditional Library		
Subject	The Book	The Paper	The Idea
Knowledge	Sacred	Formal	Informal
Memory	Persistent	Semi-persistent	Volatile
Actors	Author, Librarian	Community, Editor	Community
Dissemination	Very Slow	Fast / Slow	Very Fast
Library role	Passive	Active	Interactive

Table 1: The library paradigms

After this discussion, we will finish with our vision and a definition for the concept of a networked digital library:

A networked digital library is defined not only as an organized repository of data and information, with the traditional mission of preserving that knowledge,

but as a system with also the mission to stimulate, support and record the process of its creation.

It is now our mission to demonstrate how to turn this vision in reality.

ARQUITEC

ARQUITEC is a trial to test our vision of a networked digital library that will result in a prototype of a networked digital library for the Portuguese academic and research community.

The system will be accessible over the Internet, through a WWW interface, and will provide access to different kinds of technical documents (such as papers, reports, theses, dissertations, etc.), in any field of the knowledge. The architecture of the system is distributed, with each participating institution (universities and R&D organizations) managing its own repository (see figure 2). Based on that infrastructure, the National Library will manage an official repository of digital documents.

We intend to use ARQUITEC both as a technology demonstrator and a pilot system to develop, test and consolidate expertise in three identified issues:

Architectures of distributed digital libraries.

Procedures for management and access to the information, comprising gathering, classification, searching, retrieval and management library procedures.

Innovative services for networked digital libraries, to exploit the potential of interaction between the library and the community brought by open networks, such as the Internet.

Concerning the management of the information, the main problems will be the procedures for the remote submission of documents and their classification and search, as well as the creation and management of the official archive.

The central archive is a repository at the National Library, onto which new documents are automatically copied when they are submitted to the local repositories.

Dealing with documents from different fields of knowledge rises an important issue related with their classification and search. The key problem here is the possible integration of different metadata structures (required by the different contexts and communities) and the use of thesaurus.

We will also explore new services to be provided by the networked digital library, such as a filtering service based on the matching of the user profile and documents classification, an annotation service for documents, a collaborative catalogue and thesaurus, etc.

The library collection

ARQUITEC will provide support for a three-steps workflow in the production of information, comprising:

- **Informal document**s: a class of documents usually called grey literature (such as position papers, drafts, preprints, etc.) often useful only in the short/medium term, since it is expected that they will loose interest or they will give rise to refereed documents.

- **Refereed documents**: such as full electronic journals, papers presented in conferences or published in conventional journals, etc.

- **Formal documents:** theses, dissertations, official reports, electronic books, etc.

The increasing scholarly and scientific activity has resulted in the growth of publications rich in new interdisciplinary perspectives. That kind of contents has been raising serious classification problems for traditional libraries, where collections have been classified with catalogues usually defined by static structures. In order to deal with this dynamic classification problem, our digital library should provide users with an interactive catalog of the documents. As illustrated in figure 1, the catalog will be supported by:

- A document index.

- A multi-context and multi-lingual thesaurus (also interactive).

- The user interactions.

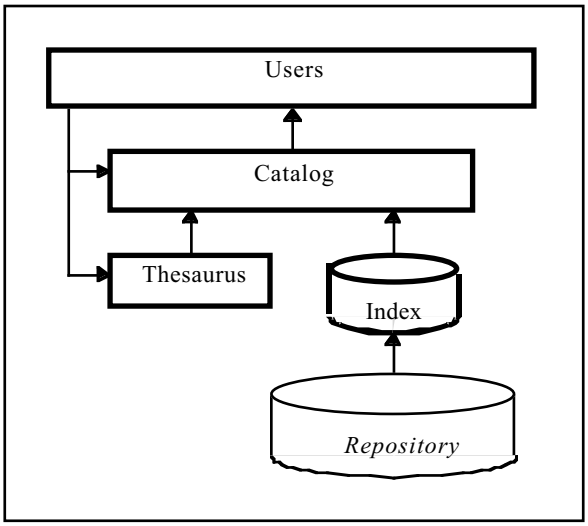

Figure 1: Interactive catalog and thesaurus

The users are able to contribute to the catalog:

- Directly, by suggesting new keywords for documents or questioning existing ones.

- Indirectly, by suggesting new relationships to the thesaurus or questioning existing ones.

For the development and interaction between the catalog and the thesaurus, experimental work was done with MCF (Gutha, 1996), a recent language for meta-content representation. For the thesaurus structure, the ISO-5964 standard was followed (ISO, 1995).

Users and services

Users can access the digital library in one of two modes: anonymous or identified. In order to register a user, the minimum required information is an electronic mail address. However, the users or the system administration can optionally provide other explicit complementary data, useful for some services (such as academic degrees, expertise fields, etc.).

An identified user has a profile, composed of the explicitly provided data and by data implicitly extracted from the history of user interactions with the system. For example, if a user retrieves a document related to a specific subject that is not in their explicit profile, this subject is implicitly added to that user's profile. Pending on explicit confirmation, this new subject will be tagged as a potential interest, which the user can easily change later.

User profiles serve three main purposes:

- Searching: for identified users, the profile is used to rank searching results, highlighting documents that match the profile (but not restricting the access to other documents).

- Filtering: the profile is also used for an information filtering service, supported by electronic mail and by the WWW interface, through which users can be notified, for example, of new documents of potential interest.

- Annotations and catalog tuning: interactive services for document annotation and catalog tuning are also provided. During an interaction with the system, any identified user may contribute also with opinions about document classification, by suggesting new keywords, questioning existing ones or by suggesting changes in thesaurus relationships. These contributions are weighed by explicit parameters of the user's profile (such as the academic degree, for example), and the results of these actions are disseminated by the electronic mailing lists related to the affected documents and subjects. This service gives users a means to interact with the library, not only to access it as an OPAC service but, in an extreme scenario, to contribute also with a new kind of meta-knowledge" that can enrich notably the library.

It is expected that the major part of the documents in the ARQUITEC digital library will be written in Portuguese or English, among other languages. Due to that, the ability to deal with more than one language will be vital for indexing and searching in documents (for example, to recognize common roots in compound words). The success of this task is one of the main targets of our project, having in mind not only ARQUITEC but also its potential application to other similar situations.

A similar problem arises with the diversity of document formats, since we don't impose a unique format. We try to support as many formats as possible, which is nice for the authors but problematic for us.

The integration of such different document formats and languages was done by the development of filters for the indexing and searching modules, rendering the format of documents transparent for the indexing and search tools.

To test solutions for those problems we have been experimenting with publicly available indexing (and searching) tools, such as Glimpse1 and Smart2. These tools have been integrated with Palavroso (Barreiro, 1993) and Correcto (Medeiros, 1995), two successful tools developed by the Natural Language Processing Group at INESC for morphologic and orthographic treatment of the Portuguese language.

Archiving and persistence

A central archive at the Portuguese National Library will be maintained, with a copy of the formal or refereed documents, after copyright has been secured from their producers. This archive will automatically harvest the new documents from the local servers, storing and cataloguing them in a central repository.

A final requirement is name persistence, especially for the documents archived at the National Library.

Depending on whether they are a serial publication or isolated books, printed documents are usually identified by ISSN or ISBN numbers. However, for digital publications such mechanism doesn't exist yet. It is usual to register CD-ROM publications with ISSN or ISBN numbers, specially if they are related to printed publications (such as the CD-ROMs distributed with magazines), but for on-line publications this is not of great help.

The publication of an on-line document is an almost instantaneous process (it requires basically the time to store and to index it in a FTP or HTTP server), and there is no expedient way to require an ISBN or ISSN number for that document compatible with this workflow. Another

important problem raised by on-line publications is that its name, or reference, should not only be an unique reference to identify that object in a specific name space, but should also provide a means to access the document (it must "say" where the object is and how to get it). This is a complex problem, globally known as URI - Uniform Resource Identifier, and its solution has been addressed by the W3C - World Wide Web Consortium3.

At present, the most commonly used form of URI is the URL - Uniform Resource Locator, but URLs have a problem: they are not persistent. If we have a document stored at a server where we need to change the structure of the stored information, the original URL of that document can become invalid, and any reference to it will originate an irritating "Error: the requested document is not valid on this server". In order to prevent that, we must ensure persistent names for stored objects, through some form of URN - Uniform Resource Name.

The problem of naming objects in a digital library was generically addressed in the CSTR project (Anderson, et. al, 1996). That work was reported in the "Kahn/Wilensky Report", from which emerged the concept of handle as an URN (Kahn & Wilensky, 1995). That concept was implemented by OCLC in the PURL - Persistent URL service.

In a few words, the PURL service is based on the existence of a highly reliable server, where it is possible to register pairs of PURLS and related URLs. In its structure, a PURL is a normal URL, with a structure like http://DNS of the PURL server…/object name.... It has a logical meaning that, when used, implies an access to the PURL server that acts as a proxy and automatically translates the logical name to the "physical" URL of the object referred to (a task performed by a simple HTTP redirect).

A PURL service, for all the persistent documents with copies archived at the National Library, will be provided in ARQUITEC. For each persistent document a PURL is automatically and registered at the central PURL server.

The global architecture

Before starting the description of the architecture of our system, we will describe some of the most paradigmatic and related projects already done in the field and whose lessons and results we used for our trial.

Related work

The CORE project started in 1991, and its purpose was to build a database of scanned journals published by the American Chemical Society (Entlich et. al, 1995).

By the end of 1994 they had a database of more than 400,000 pages of full text and graphics (in magnetic tapes and CD-ROM). The text was converted to ASCII and marked-up with SGML (Standard General Markup Language), the database being accessible with dedicated X-Windows interfaces. The other major contributors of this project were the Cornell University, OCLC, Bellcore and Chemical Abstract Service.

The users accepted the results of the CORE project very well, but another conclusion was also that "the task of building and maintaining electronic journal databases remains formidable."

A contemporary and also ambitious initiative was the TULIP project, started in March 1991 and concluded in the end of 1995 (Elsevier, 1996). It was sponsored by Elsevier Science, and involved nine universities in the USA (C.M.U., Cornell, Georgia Institute of Technology, MIT, Univ. of California, Univ. of Michigan, Univ. of Tennessee, Univ. of Washington, and Virginia Polytechnic and State Univ.).

The main goal of the project was to research and test systems for networked delivery and use of scanned journals. Elsevier contributed with the scanned page images, OCR generated text and bibliographic data from 43 engineering and materials science journals. The universities provided solutions to deliver these journals in electronic form to their users. The research focus was on technical issues, user behavior and organizational and economic problems.

When the project TULIP started, the Internet was already a reality, but the Web was still in an embryonic state. Due to that, the delivery technology was based on dedicated graphical clients for X-Windows, MS-Windows and Apple Macintosh, besides alphanumeric clients for mainframe terminals. But soon the maintenance costs were evident, and the project shifted to WWW technology when its advantages and maturity became recognized.

In its final conclusions, the project pointed out that the transition from conventional to digital libraries (defined here as libraries with full digital contents), will take much longer and cost more than commonly thought, mainly due to network bandwidth and storage limitations.

However, and as it was also pointed out by the CORE project, we think that this conclusion can not be dissociated from the approach taken: to scan the original material. For example, it was estimated in TULIP that a typical journal issue, with 20 articles and 200 pages, requires approximately 17 Mbytes of storage, with 16 Mbytes for the scanned pages (in TIFF format). By comparison, the ASCII information resulting from the OCR process requires only 800 Kbytes and the indexing and bibliographic information (in SGML format) requires about 200 Kbytes.

More pragmatic approaches were taken in a series of projects in the Computer Science Reports area. Some of the most representative were UCSTRI - Unified Computer Science Technical Report Index (VanHeyningen, 1994), NTRS - NASA Technical Report Server (Nelson, et. al, 1994), WATERS - Wide Area Technical Report Service (French, et. al, 1995) and CSTR - Computer Science Technical Reports. A common goal of those projects has been easy installation and maintenance of the server sites and support for heterogeneous collections. The idea has been not only to provide scanned versions of printed documents, but also to take advantage of the fact that today it is normal to produce, in the source, those documents already in digital formats (such as ASCII, MS-Word, PDF, HTML, etc.).

In April 1995, WATERS and CSTR projects joined efforts and conceived a new service: NCSTRL - Networked Computer Science Technical Reports Library (Davis, 1995). NCSTRL is a network of servers providing three kinds of services: repository, indexing and user interface. Currently NCSTRL is a worldwide service, with repositories installed in over 60 universities and research centers across the world. NDLTD, a more recent project in the USA, aims to extend that base to provide a generic national digital library of theses and dissertations (Fox, et. al, 1996).

DIENST and NCSTRL
INESC has been experimenting with the NCSTRL technology since middle 1996. We were impressed by its capabilities as a potential framework for future work, especially its open architecture model and its ability to handle documents in several formats. Therefore we decided to use it as the core technology for ARQUITEC. In figure 2 we present the main blocks of that architecture.

The DIENST technology was the main contribution of project CSTR for the NCSTRL initiative (Davis & Lagoze, 1994). The NCSTRL architecture is based on a network of DIENST servers (referred to as S), each one managing a repository of documents (R) the respective index (I) and user interface (UI). The user interface is implemented in HTML, provided through an HTTP server (the DIENST server is written in PERL and its interface to the HTTP server uses CGI). A user can access any server from any user interface, since user searches are always performed in all the indexes.

Optionally, the repositories can be accessed via lite servers (L), the main contribution of project WATERS for NCSTRL. In this case each site only has to provide a metadata description file (M) and have its documents accessible by FTP or HTTP. The lite server converts that metadata to the DIENST format, indexes it, and provides normal DIENST interfaces for the users and for the other DIENST servers. In the specific case of the NCSTRL service, it has only one central server for all the registered lite repositories.

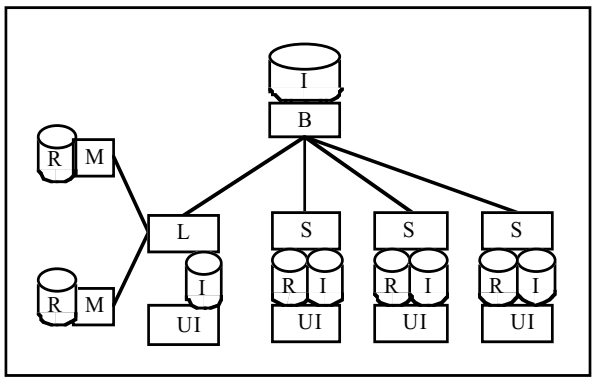

Figure 2: The NCSTRL architecture.

A backup server (B) can maintain a copy of all the indexes, which is useful if one of the servers becomes inaccessible. In that case users will not be able to perform retrievals, but at least they will be able to search and find references to the desired documents.

Finally, our architecture
The architecture of ARQUITEC is distributed, with local nodes managing the local repositories at the universities and research institutes, but all the collections are freely accessible for search from any node. The core of ARQUITEC is based on a modified and extended version of DIENST 4.0. The required modifications occurred at the three modules of NCSTRL, corresponding to three different tasks of ARQUITEC: replacement of the indexing and searching tool, modification of the repository management and modification of the interface.

The original DIENST indexing and search tool had to be replaced by a more powerful catalog, as described. The new requirements implied modifications at the NCSTRL repository interface level, in order to perform full text indexing of as many document formats as possible (such as ASCII, Postscript, MS-Word, etc.), as well as in different languages.

Concerning the management of information, the main generic problems were the procedures for submission of the documents, their classification and search, as well as the creation and management of the central archive.

The submission of documents can be done remotely, with the user authenticated by username and password (stronger security and authentication issues, for which we recognize the importance, were not addressed for now). The submission process starts by the filling and submission of

registration forms, by WWW. Users will be required to provide the location of the original document at an FTP or HTTP server. After that a confirmation procedure takes place: an electronic mail message is sent to the user and the system waits for a reply. After successful confirmation, the document is then retrieved, registered and added to the catalog.

The core of the NCSTRL system was also modified in order to allow the automatic management of the official central archive. In practice this means that the central host, at the National Library, automatically gathers all new persistent indexed documents into a central repository. That repository is used as an official archive, which is especially important for theses and dissertations. It also serves as a mirror repository to provide global fault tolerance.

The NCSTRL user interface was modified in order to support all the described requirements, new functions and services. The modifications were done essentially in the submission of documents (that can now be done remotely), as also in the support of the search task. All the interface components were redesigned to support multi-lingual access (Portuguese and English in the first release).

Finally, a directory for the registered users was added to the system. It is a distributed directory based in the X.500 model, with an LDAP interface (Yeong et. al, 1995).

Future work and open issues

Medium term work will be concerned with the integration of other spaces, accessible by new interfaces at lite DIENST servers. Examples will be interfaces for Z39.50 servers4, useful for the integration of OPAC systems such as the catalogs of conventional libraries, and HARVEST5 brokers, useful for the support of informal publications and other similar material such as mailing lists, source code, etc.

Examples of other identified research issues requiring our attention in the medium/long term are:

- Document structuring: research will be done on using SGML and other alternative solutions for structuring the information objects (a specially interesting issue to be applied not only for the original documents but also to represent the associated annotations);

- Natural language: trials will be done in the classification and search of documents with natural language techniques, with a special concern for the Portuguese language;

- Authentication and certification authorities: the requirements for authentication and certification authorities, for both the documents and users, will be addressed in medium term;

- Legal issues: among generic problems, such as how to assign and observe other properties of the documents (such as terms and conditions and other copyright problems), examples of new open interesting problems in this field are the legal implications of the new objects, composed by an original document and a list of annotations (or just the legal implications of an annotation);

- Long term preservation: how will the official repository survive the evolution of the hardware and software, such as storage technology, operating systems, document formats, viewers, etc.?

References

Anderson, G.; Lasher, R.; Reich, V. (1996). The Computer Science Technical Report (CS-TR) Project: A Pioneering Digital Library Project Viewed from a Library Perspective. The Public-Access Computer Systems Review 7, No 2, 1996. Available at http://info.lib.uh.edu/pr/v7/n2/ande7n2.html

Barreiro, A.; Pereira, M., J.; Santos, D. (1993). Linguistic options and criteria in the development of Palavroso, a computational system for the morphological description of Portuguese (in Portuguese). INESC Report No. RT/54-93, December 1993.

Davis, J. R. (1995). Creating a Networked Computer Science Technical Report Library. D-Lib Magazine, September 1995. Available at http://www.dlib.org/dlib/september95/09davis.html

Davis, J. R.; Lagoze, C. (1994). A protocol and server for a distributed digital technical report library. Technical Report TR94-1418, Computer Science Department, Cornell University, 1994.

Elsevier Science (1996). TULIP Final Report. Elsevier Science Edition. Available at http://www.elsevier.nl/locate/tulip.

Entlich, R.; Garson, L.; Lesk, M.; Normore, L.; Olsen, J.; Weibel, S. (1995). Making a Digital Library: The Chemistry Online Retrieval Experiment. Communications of the ACM, April 1995, Vol. 38, No. 4, 54.

Fox, E. A.; Eaton, J. L.; McMillan, G.; Kipp, N. A.; Weiss, L.; Arce, E.; Guyer, S. (1996). National Digital Library of Theses and Dissertations. D-Lib Magazine, September 1996. Available at http://www.dlib.org/dlib/september96/theses/09fox.html

French, J. C.; Fox, E. A.; Maly, K. (1995). Wide Area Technical Report Service: Technical Reports Online. Communications of the ACM, April 1995, Vol. 38, No. 4, 45.

Gutha, R. V. (1996). Meta-Content Format. Apple Computer. Available at http://mcf.research.apple.com/hs/mcf.html

Harnad, S. (1990). Scholarly Skywriting and the Prepublication Continuum of Scientific Inquiry. Psychological Science 1, 342 - 343.

Harnad, S. (1991). Post-Gutenberg Galaxy: The Fourth Revolution in the Means of Production of Knowledge. Public-Access Computer Systems Review 2, No 1, 39-53. Available at http://info.lib.uh.edu/pr/v2/n1/harnad.2n1.

Harnad, S. (1995). The PostGutemberg Galaxy: How to Get There form Here. The Information Society 11(4), 285-291.

ISO - International Organization for Standardization (1995). ISO-5964: Documentation Guidelines for the establishment and development of multilingual thesaurus. Geneva, 1985.

Kahn, R.; Wilensky, R. (1995). A Framework for a Distributed Digital Object Services. Available at http://WWW.CNRI.Reston.VA.US/home/cstr/arch/k-w.html

Medeiros, J., C.,(1995). Processamento Morfológico e Correccao Ortográfica do Português. Master Thesis, Instituto Superior Técnico - Universidade Técnica de Lisboa, Lisboa.

Negroponte, N. (1996). Ser Digital. Editorial Caminho (Portuguese Edition of the original title “Being Digital”, 1995).

Nelson, M. L.; Gottlich, G. L.; Bianco, D. J.; Paulson, S. P.; Binkley, R. L.; Kellog, Y. D.; Beaumont, C. J.; Schmunk, R. B.; Kurtz, M. J.; Accomazzi, A.; Syed, O. (1994). The NASA Technical Report Server. Internet Research: Electronic Network Applications and Policy, Vol. 5, No 2, 25-36.

O’Reilly, T. (1996). Publishing Models for Internet Commerce. Communications of the ACM, June 1996, Vol. 39, No 6, 79-86.

Okerson, A. S.; O’Donnell, J. (1995). Scholarly Journals at the Crossroads: A Subversive Proposal for Electronic Publishing. Association of Research Libraries, June 1995.

VanHeyningen, M. (1994). The Unified Computer Science Technical Report Index: Lessons in Indexing Diverse Resources. Second International World Wide Web Conference, WWW’94 Oct. 94, 535-543.

Yeong, W.; Howes, T.; Kille, S. (1995). RFC 1777: Lightweight Directory Access Protocol. IETF Network Working Group. Available at http://www.umich.edu/~rsug/ldap/doc/rfc/rfc1777.txt.

1 http://glimpse.cs.arizona.edu

2 ftp://ftp.cs.cornell.edu/pub/smart

3 http://www.w3.org/WWW/Addressing/Addressing.html

4 http://lcweb.loc.gov/z3950/agency

5 http://harvest.transarc.com

* Paper presented at IASSIST/IFDO ‘97, Odense, Denmark, May 6-9,1997.

José Luis Borbinha (Jose.Borbinha@inesc.pt) IST - Technical Superior Institute (Lisbon Technical University) Department of Electrical and Computers Engineering
José Delgado (Jose.Delgado@inesc.pt) INESC - Institute for Systems and Computer Engineering Telematics Systems and Services Group