July 2005

From the Physical Reality to the Virtual Reality in the Library Environment

Markos Dendrinos
Technological Educational Institute of Athens

Library Philosophy and Practice Vol. 7, No. 2 (Spring 2005)
(libr.unl.edu:2000/LPP/lppv7n2.htm)
ISSN 1522-0222

From the Physical Reality to the Virtual Reality in the Library Environment

Markos Dendrinos
Associate Professor of Computer Science
Technological Educational Institution of Athens (TEI-A)
Dept. of Librarianship and Information Systems
Ag. Spyridonos, 122 10
Egaleo, Greece

Introduction

The book in various forms (handwritten papyri, parchments, paper based etc.) remained for a long time the main carrier of assembled information. Nowadays, after the great revolution in the information and communication technologies (ICT), an extension of the notion of the book has taken place, leading to the notion of the information resource. In such a wide frame any type of resource is included, since the only criterion for the consideration of a module as a resource is its informative character. The library as a building has been transformed to the library as an environment of electronic services established on a computer server or a network of cooperating servers. The client is not obliged to go to a certain place for searching and retrieving his information, since the information is coded in a binary form being available to any person linked locally or through the Internet to the specific server.

Copies of a book are real entities stored somewhere in the selves of a library and each of them can be taken and read by a user. In the virtual environment, the notion of the "copy" has been eliminated, since the information is omnipresent to any person equipped with communication link, communication protocols, and decoding software compatible with the one used by the stored information.

Types of resources have also been expanded. Print is no longer the only way of representing information. Texts, still images, video, audio (sounds, music and speech), and any combination of types are just some of the information carriers in the multimedia era. Resources are digital from the beginning or they are the digitized version of corresponding analog forms. An ongoing transition to digital repositories through immense "retrodigitization" programs is taking place. (Thaller 2001)

Electronic Resources

Electronic resources are also available on the Internet (commonly indicated as Networked Information Resources or NIRs). Although access to most NIRs is currently uncontrolled and free, controlled access (e.g., through user registration and passwords), and fee-based access (based on networked payment mechanisms) are options, which allow the addition of commercial and copyrighted materials to the range of networked information resources (Mackenzie & Wierck 1996, p. 9)

Electronic documents differ radically from printed ones. Printed documents are typically distinct entities of a static, unchanging nature. Initially, electronic documents were just imitations of printed documents in electronic form. Recently, they have become more and more compound and dynamic. Compound means that the document can consist of various distributed "information objects," discrete components, each of which can be at a different physical location. It can also include metadata (e.g. author, revision history, etc.) and links to external multimedia modules (images, video, audio, graphics, etc.) An electronic document can be nothing more than a set of pointers to a number of components. Dynamic means that any component of the document can change, giving the document a temporary nature (Mackenzie & Wierck 1996, p. 10-11).

It is obvious that the up-to-date nature of a NIR is its strength but also its weakness. It is exactly what someone needs in a rapidly changing world and at the same time is something that cannot be used as a reliable and constant citation, since the reliability of a resource can be assured only by its maintainer.

Another critical factor in the effective use of the resources is the foundation of a common ground of standards, which is absolutely necessary for the improvement of their interoperability. The concept of interoperability includes wide usefulness (re-usefulness), portability (across networks, systems, and organisations) and longevity (portability across time). The key to the interoperability of content is consistency, and consistency is achieved through the use of standards (Gill & Miller 2002, p. 3).

Digital Libraries

The working environment of electronic resource collections is still called a library. Such a library is called digital or virtual, namely, "a library that provides access to a collection of distributed information in electronic format through pointers provided locally … or a collection of digital objects housed in the same place, virtual or physical" (Keyes 1999, p. 180). A digital library is also described as "a zone of convergence where librarians, computer scientists, electrical engineers, cognitive scientists, cultural anthropologists, organizational theorists and sundry others are forging a lingua franca for better understanding the nature of distributed information systems and knowledge access" (Cronin 1998, p. 43).

The digital libraries sector is a rapidly developing area, combining many areas of research and technology. From a database or information retrieval perspective, digital libraries may be seen as a form of federated databases. From a hypertext perspective, digital libraries could seem like a particular application of hypertext technology. From a wide-area information service perspective, they could appear to be one use of the World Wide Web. From a library science perspective, they might be incorporated in the trend toward library automation (Nuernberg et al. 1995, p. 1-2). Obviously, a digital library is not confined to any of those fields but it must be considered a union of subfields with great added value, because one can participate more actively in searching and retrieving information in a relatively controlled environment. A digital library remains a library incorporating the existing tradition of physical libraries, but at the same time offering new capabilities, impossible in the previous conventional environment.

Following this view, we can make use of the physical library elements as a starting point for discussing the elements and domains of digital libraries, considering three broad classes of library elements: data, metadata, and processes (Nuernberg et al. 1995, p. 2).

Access to Networked Resources

At the present, libraries tend to use the existing cataloguing rules for the bibliographic control and access to networked resources. This is a natural decision, since it requires relatively little change in familiar practices and procedures and a low investment in systems. In the future this approach will not be satisfactory, since it will provide little support for the complicated task of networked cataloguing and offer the user less functionality and ease of use. Thus, there is an urgent need for a new generation of library systems, designed to cope with the requirements of integrated networked services (Mackenzie & Wierck 1996, p. 71).

It is a question whether automated digital libraries can rival conventional libraries today. Let us see the contrast between web search engines and conventional abstracting and indexing services or library catalogs. The selection of material indexed by a web search engine is automatic based on arbitrary considerations, there is no authority control, and the content can contain many dead links. On the other hand, cataloguing is expensive, while web indexing is cheap. The leading web search engines index several hundred million web pages every month, more than the total number of MARC records that have ever been created (Arms 2000, p. 2-3). We cannot claim the superiority of the one system to the other, since it depends on the priorities set by the user.

The new kind of resources demand a new set of skills, and even an entirely new relationship with publishers. During the negotiation process between the librarian and the publisher, the librarian must agree to certain restrictions on photocopying or distribution of electronic materials. The library is responsible for maintaining the awareness in all employees about copyright issues (Orick 2000, p. 316-317).

The number of electronic resources is increasing rapidly. Libraries are paying more attention to their ability to get access to these resources than to archiving and protecting print materials for users in the future. Unlike a print book or journal, electronic resources cannot be considered permanent additions to a collection. It is not the product itself that is purchased but access to the product. The publisher and not the librarian dictates how much access will be provided, which issues will be available, and how much access will cost (Orick 2000, p. 318)

Data, Metadata, and Processes

Data are library materials, metadata is information about them, and processes are functions performed over library elements. We can find digital library data corresponding to physical objects. Texts, still images, videos , and recorded audio can be digitized and stored in the memory of the server. When translating a book into a series of digital images, all of its information is retained except perhaps some exceedingly specific details. Conversion of a printed book into ASCII text to be used in full text retrieval is followed by the need for considerable

decision-making and labor regarding its final form: various fonts and letter sizes, graphics and images will be lost, since only the basic text is preserved.

A classification scheme such as Library of Congress Classification is an example of metadata. A classification scheme does not have any physical reality itself, but its application is sometimes constrained by the physicality of the objects it classifies (Nuernberg et al. 1995, p. 5). Sometimes it is used as a guide to the physical location of data in a library, since placing similarly-classified objects in physical proximity can aid patrons in locating data. We can see here the superiority of the virtual library, whose abstract space can be used for multiple classifications of a book, a fact impossible in the frame of a physical library.

An example of a physical library process a librarian helping a patron find something for which he or she has incomplete information. This process is necessary in an integrated virtual environment, since digital library patrons cannot rely on being near someone who can help them. While the digital library starts at the physical level, it continues with elements and services absent from the physical library environment. Hypertext and active computational objects are examples of new library data. Multiple classification schemes are examples of new library metadata. Lastly, sophisticated human–machine interfaces used for more efficient and user-friendly communication and adapted also to users with special needs are an example of new automated library processes.

Digital Libraries Initiative

The Digital Libraries Initiative (DLI) is a major US governement initiative funded by the National Science Foundation (NSF), the Defense Advanced Research Projects Agency (DARPA), and the National Aeronautics and Space Administration (NASA). Projects of the DLI pursued deep semantic interoperability, uniting heterogeneous items in a single uniform federated source (Schatz & Chen 1999, p. 48-49). In the net of the 21st century there will be a billion repositories distributed over the world, each belonging to a certain community. Semantic indexes operated through scalable semantics and concept switching will be effective in bridging the gaps between the specialised terminologies of the various communitities (Schatz, et al. 1995, p. 55). The arbitrarily hyper-linked Internet will be gradually transformed into a semantic Interspace. For the extension of web searching beyond full-text retrieval, document structure in the short run, and document semantics in the long run, are required (Wactlar et al. 1999, p. 66-73).

The expanded resource type in the digital library environment has had an unavoidable impact on the query and retrieval processes. The Informedia DLI project at Carnegie Mellon University led to a terabyte digital video library. Automatically derived descriptors were used for indexing, segmenting, and accessing the library contents. Speech recognition, image compression, face and colour detection, and video optical character recognition (VOCR) combined in a very sophisticated manner resulted in an effective multimedia retrieval mechanism (Paepcke et al. 1999, p. 86). In general, the query techniques used in the digital library environment will be the result of an optimal integration of traditional library facilities and current web-based approaches (Capurro 1996, 256-270).

Forms and Information

The virtual nature of resources could relate them to the ancient concept of Platonic *eidos*/idea. Capurro gives a very interesting interpretation of the concept of information through its etymological connection to the concept of form. Information in this sense is the process of forming a piece of matter or, metaphorically, human knowledge (Capurro 1996, 256-270). Shannon and Weaver in their mathematical theory of communication establish a neutral, independent of the human comprehension, meaning of information content. Carl Friedrich von Weizsaker relates this neutral substance, which is neither matter nor energy to the Platonic *eidos* and the Aristotelian form (Capurro 1996). Therefore, information content is meant as an autonomous substance, which drives every kind of process, a pattern archetypon for the changing or in-forming something (Capurro 1991).

We can consider the virtual resource as the information content beyond and before any information process. The virtual resource coded in a machine language and stored in a machine memory can be transformed to a human sensed resource through the process of information . Something existing but not sensed by the human factor can take many forms, either on the screen or printed, conveying meaning to the human. This can be considered as analogous to the platonic *eidos*, something not accessible by the human senses, which can take a multitude of material forms, comprehensible by some species of material life, including human.

Conclusion

Everything seems ideal in that virtual information world. But in reality there is an inequity of access to the information content either from home or from the library, since the many libraries lag far behind in providing their patrons with PCs with Internet access. The Library Bill of Rights says : “electronic information flows across boundaries and barriers despite attempts by individuals, governments and private entities to channel or control it. Even so, many people, for reasons of technology, infrastructure, or socioeconomic status do not have access to electronic information” (Orick 2000, p. 317).

Access to the platonic *eidos* is not a matter of nationality, religion, gender, or socioeconomic status. All members of the information science community must extend access to virtual resources, the common fortune of the human kind, to all the layers of individuals and communities.

Works Cited

Arms, W. (2000). Automated Digital Libraries. *D-Lib Magazine* 6.7/8 Jul/Aug. pp. 2-3.

Capurro, R. (1991). *Foundations of Information Science, Review and Perspectives, International Conference on Conceptions of Library and Information Science, University of Tampere, Tampere, Finland.* Also available at www.capurro.de/tampere91.htm

Capurro, R. (1996). *On the genealogy of information, international conference: information, new question to a multidisciplinary concept*. German Society for System Research. Berlin: Akademic Verlag. pp. 259-270, also available at www.capurro.de/cottinf.htm

Cronin, B. (1998). Information Professionals in the Digital Age. *Intl. Inform. & Libr. Rev*. 30, p.43.

Gill, T., Miller, P. (2002). Re-inventing the Wheel? Standards, Interoperability and Digital Cultural Content. *D-Lib Magazine*, 8.1 (Jan.), p.3

Keyes, M. (1999). The Evolving Virtual Library: A Vision through a Glass, Darkly. In L.M. Saundres (Ed.) *The evolving virtual library II*. Medford: Information Today, Inc. p. 180.

Mackenzie Owen, J.S. , Wierck, A. (1996). *Knowledge models for networked library services – libraries in the information society*. European Commission, DG XIII-E/4. p. 9

Nuernberg, P. Furuta, R. Leggett, J., Marshall, C. C., Shipman III, F.M. (1995). *Digital Libraries '95, "Digital Libraries : Issues and Architectures", The Second Annual Conference on the Theory and Practice of Digital Libraries, June 11-13, 1995, Austin, Texas, USA*, also available in csdl.tamu.edu/DL95/papers/ nuernberg/ nuernberg.html, pp. 1-2.

Orick, J. T. (2000). The Virtual Library: Changing Roles and Ethical Challenges for Librarians. Intl. *Inform. & Libr. Rev.* 32, pp. 316-317.

Paepcke, A., Wang Baldonado, M. K., Chang, C. K., Cousins, S., Garcia-Molina, H. (1999). Using Distributed Objects to Build the Stanford Digital Library Infobus. *Computer*, IEEE, (Feb.), p. 86.

Schatz, B., Chen, H. (1999) Digital Libraries Technological Advances and Social Impacts, *Computer*, IEEE, Feb., pp. 48-49.

Schatz, B., Mischo, W., Cole, T., Bishop, A., Harum, S., Johnson, E., Neumann, L., Chen, H. (1999). Federated Search of Scientific Literature. *Computer*, IEEE, Feb., p. 55.

Thaller, M. (2001) From the Digitized to the Digital Library. *D-Lib Magazine* 7.2 (Feb.)

Wactlar, H., Christel, M., Gong. Y., Hauptmann, A. (1999). Lessons Learned from Building a Terabyte Digital Video Library. *Computer*, IEEE, (Feb.), pp. 66-73.