63RD IFLA Council and General Conference

UDT Core Programme Workshop: September 4, 1997

METADATA, LIBRARIES AND LIBRARIANSHIP
(speech notes)

Terry Kuny
XIST Inc. / UDT Core Programme
email: terry.kuny@xist.com

Last revised: August 27, 1997

For those of you who have read earlier the March 1997 Scientific American article on searching the Internet, the news is now out: The Internet is not the Digital Library.

To those of you who have tried an Internet search engine and retrieved 50,000 results, or who have spent countless hours honing your skills at winnowing nuggets of valuable information from the vast dross that clutters the network, then this news comes as no surprise. The single most important challenge in our networked future is "information overload"—how to find and manage a plethora of information.

As Clifford Lynch so simply states in his article, if the Internet is to continue to grow and thrive as a means of communication, then "something very much like traditional library services will be needed to organize, access and preserve networked information."

Dr. Lynch's other important point is that we must combine the skills of the librarian and the computer scientist to help organize the current anarchy of the Internet. Today's workshop is an effort to bring different communities together for the common purpose of addressing the problem of how to simply describe information on the Internet. This is the subject of today: "metadata" - information about information.

By providing information about information, we can build retrieval systems that can more accurately and efficient locate resources that are relevant to users. Virtually all "knowbot" or "intelligent agent" tools of the future will look for some form of metadata, or will create some form of metadata, to aid in information retrieval. The big "Star Trek" ("Computer: give me ….") databases like Alta Vista will not scale to provide the rich retrieval in a multi-billion heterogeneous document index without the use of metadata. Providing structured information and structured searches will assist in local document management and retrieval, and will contribute to enhanced retrieval in the larger Internet community.

It strikes me that the efforts that you will hear about today to create "metadata" standards for the Internet, brings us full circle to the earliest days of librarianship. The same challenges arose in developing rules for cataloguing. AACR2 was not passed to librarianship as laws sent from the heavens. There is a small but vital history of librarians coming together to argue about and agree upon rules which allow us to make sense of the anarchy of the printed word. This is a substantial achievement of our profession and the organization of a library is dependent upon those rules.

So today, in our electronic environment, we face the same problem, how to describe information. We are creating a new history of cataloguing. It is a history of individuals such as yourselves, coming together to learn and argue and finally, to agree upon rules which will allow the common person to make greater sense of the networked world.

It has taken over one hundred years for cataloguing standards to become well-established and codified. Given the pace of change in technologies, I feel confident in predicting that it will not take one hundred years for the same to occur in the networked world. For without these standards, all the potential of the Internet and the possibilities of digital libraries will remain unrealized.

It is a particular irony of the Internet, often portrayed in the popular press as a digital library, requires that everyone become more like librarians. Internet users who hope to find information are now required to learn Boolean searching, to formulate complex queries, and to become better aware of different information resources. Similarly, people wanting to provide information that can be effectively managed or retrieved on the network must develop some of the skills of a cataloger.

ROLES FOR LIBRARIES AND LIBRARIANS

So what are some of the new roles for librarians and libraries?

- Selection, Evaluation and Description of Networked Information.

  We need to manage information overload for users. We need to become sense-makers for our users. Librarians are intelligent agents and can help build intelligent tools for users.

- Coding Metadata Records.

  We need to begin working with metadata and using it to describe networked information. By doing this, we can help identify problems and suggest solutions. There are many different forms of metadata. Stuart will be talking about the Dublin Core, a kind of lowest common denominator of metadata formats. In it design, the Dublin Core metadata standard is directed at addressing some of the problems of Internet-based information. Still other user communities will require richer descriptive information and they may use the Government Information Locator Service metadata model or the Text Encoding Initiative headers. There will be many kinds of metadata, each appropriate to its use.

- Standards Setting.

  Many librarians have a rich knowledge of the problems and pitfalls of cataloguing and for resource description. There are a number of metadata standards activities where this experience would be helpful. The

profession needs to increase its knowledge about metadata activities and the best way to do this is through participation.

- Training.

  Training in the use of metadata, Examples include good practices for coding; document management; preservation of metadata; and search techniques. A future meeting of the Dublin Core will probably be devoted to user training and guideline creation. This type of activity is a good place for librarians to become involved.

- Advocacy.

  Building an awareness of the problems in resource discovery and bringing this onto the information policy agenda of funding agencies is vital. As libraries already know, cataloguing is expensive. Although there will be many automated means for the creating of metadata, the development and operation of networked resource discovery services will be costly. I believe that IFLA can be a catalyst in promoting the use of metadata on the Internet and I expect that the UDT Core Programme and Section on Information Technology will focus some of their efforts on metadata activities in the coming years.

- Document Repositories.

  There is something of a debate over whether libraries should “link” or “collect.” Building document repositories is one mechanism that libraries can ensure that materials are available to their users. This has the added benefit of getting our hands dirty by providing librarians with first-hand experience in managing electronic information. The digital library research agenda is a long one and there is important research to be done in the management of digital collections that libraries can contribute to.

- Registries (PURLs; metadata; naming).

  Among the key information infrastructure requirements for networked resource discovery and retrieval are the availability of agencies which act to provide particular kinds of public goods to the network community. The management of domain names by Internic or the MIME registry at the Internet Assigned Names Authorty are examples of such services. These services are key to the smooth operation of the Internet.

  There are similar services that will be required to help build an efficient infrastructure for information retrieval. Among these will be the provision of registry services that are universal locators—persistent uniform resource locators (PURLs) which can potentially last indefinitely. As there are more and more digital objects available on the network, building persistence in location will be a critical development. As there will also be different metadata schemas being used for different purposes, building registries to assist in semantic interoperability, that is, to share information

about what the contents of various schemas mean. Such registries can also provide information about associated interchange formats.

CONCLUSION

Our digitally enhanced future still requires the intellectual apparatus of the librarian, even as the physical walls of the library become less important. Of course, we can expect great gifts from our colleagues in computer and information science, especially in developing tools and standards to make our work more efficient. But I feel comfortable in believing, along with Clifford Lynch and many other librarians, that there will be no digital library without good digital librarians and their continuing commitment to creating islands of information stability in a data storm.