INTRODUCTION

The world is inundated with information. Since the appearance of the Web in the mid-1990s, information has become available to anyone at the click of a button.

In 2000, the Web had more than 4 billion public pages and an additional 550 billion connected documents in the deep Web. By 2000, the Web was written in more than 220 languages, 78% of which were in English with 7 million pages added daily. The average life span of a Web page was only 44 day, and 44% of Web sites available in 1998 had disappeared by 1999. (Lyman and Varian 2000)

Despite the popularity of the Web, many concerns exist regarding the discovery, organization, retrieval, and description of information in this new digital universe. The familiar library challenges of describing and organizing information in the print world multiply in the virtual world. For instance, the entire concept of seriality and serial publications in cyberspace requires new definitions and approaches. Web pages are updated constantly, in terms of redesign, graphics, and content. Do those updates make Web pages serials?

And what about online newsletters and electronic versions of print serials? In the print environment, librarians have always had to address similar problems regarding reissued, revised, and updated government documents and items, but the Internet has only magnified these challenges. Other concerns include authenticity and reliability of Internet information (especially in the areas of health and medicine), filtering, privacy, and copyright.

The explosion of electronic projects that focus on the digitization and organization of an institution's unique resources has revitalized interest in special and local collections around the world. Suddenly, money is available to anyone who can spark interest in his or her unique resources by identifying grant sources and writing intelligent grant proposals.

The swell of digital libraries and digitization projects has reintroduced the age-old questions of how to manage and describe information. Not only are the challenges of information organization and description in the print environment still being discussed, but now the Internet and the electronic environment have magnified the problems.

If a document is available in both the print and electronic environment, for instance, but the electronic document provides hyperlinks to references and has been given a slightly different title on the Web page, are the two documents still the same? When a Web page has been substantially revised or redesigned, does this change necessitate a new bibliographic record? The debate on whether to organize information by format/container (print, video, audio, CD, DVD, and so on) or by content (focusing on the information itself rather than how it is presented) has again resurfaced in the professional cataloging community.

Libraries and related information organizations such as museums and archives have always been the repositories of human knowledge and information. These information communities have been developing standards that are uniquely related to their own information and patron needs in the digital era.

The **deep Web** is that part of the Web that cannot be indexed by search engine spiders and bots, such as PDFs, Web sites that incorporate frames structure, and databases.

A **PDF** is a portable document file.

Metadata helps information providers handle the challenges of organizing information for access. Metadata is a relatively new buzzword in the digital universe, but its purpose and meaning have been around as long as human-kind. Its most familiar emanation is known as cataloging.

For librarians, metadata is just a fancy term for the organization and description of information in its many forms. Organization and preservation are indeed the main challenges and concerns for those confronting the production and efficient search and recall of digital information, as well as its future preservation and usefulness.

Many organizations have created proprietary sets of metadata standards. The current emphasis of the metadata community is to create interoperability between and among these many standards, as well as to begin the process to narrow and refine the number of standards to a few acceptable and agreeable general ones.

This report is a guide to current metadata standards and their application in whatever environment you are working in. Because of the vast number and variety of metadata standards under construction, only the major standards are included in this report. Besides discussing these standards and their histories, this report examines:

- Which metadata is suitable for certain libraries

- Linking initiatives and how they relate to metadata

- How to use metadata to build an enriched library catalog

- How metadata assists in natural language recognition technology

The importance of interoperability and crosswalks, as well as the uses of metadata in technical services, cataloging, and digital projects are included.

Due to the fast-paced developments in this area, this report can only attempt to document the current state of these standards at the time of publication. The official Web sites and major projects that are implementing these metadata standards are listed; they are the best ways to maintain currency and accuracy of information on metadata standards.

Crosswalks are human-designed indexes that attempt to match and pair different tags and elements in the various metadata schemes, so that Web and digital resource creators understand their relationships and can assist in interoperability for retrieval and access.