DIGITAL EVOLUTION: CHANGING ROLES AND CHALLENGES FOR ARCHIVISTS IN THE AGE OF GLOBAL NETWORKING

LAURIE B. CRUM

ABSTRACT: The combination of the evolving nature of the Internet and new document types and information that are being disseminated throughout the World Wide Web pose significant challenges for any professional engaged in the documentation and study of organizations and human culture. This article serves as an exploratory platform for archival issues related to global computer networking as exemplified by the Internet. A discussion of the parallels and differences of the evolution of the book and hypermedia World Wide Web documents highlights a number of important challenges for archivists. The article also offers dynamic, organic definitions of the Internet and hypermedia World Wide Web documents. The author examines important cultural implications and concerns for professionals engaged in the use and documentation of new forms of digital media.

Introduction

Rapid developments in computer networking technology and digital multimedia, world-wide competition, economic trends, and the desire for faster and broader dissemination of information are causing many information professionals, including archivists, to consider the implications for their specific professional roles. Traditionally, when society and its organizations have implemented new technologies that create new forms of documentation, archivists have coped with the management issues raised by the products of these technologies long after they have come into use. This approach is no longer viable. With the advent of the "information superhighway" and increasing worldwide access to information through the Internet, archivists have not only an opportunity, but an obligation to position themselves at the forefront of these developments and meet the issues and challenges raised by the products of digital networking and hypermedia documents. The purpose of this article is to begin professional discussion on these issues and challenges. In doing so, the article makes some observations on the organic nature, and offers preliminary definitions, of the Internet and its contents that may assist archivists in their considerations of identification and management of digital documents created by their own institutions or that fall within the scope of their collecting mandates. This discussion

focuses in particular on uses of hypermedia World Wide Web documents (viewed through software such as NCSA Mosaic and Netscape), being used by organizations, groups, and individuals to prepare, distribute, and annotate multimedia documents or "HomePages" for informational and transactional purposes. This article also compares the current digital evolution to the evolution of the book, and develops an adaptive definition of digital documents that might be used as a framework for many of the issues archivists face related to the Internet.

The Internet

Traditionally, records are considered to be static. (For example, a written memorandum from the president to the vice-president of a company is considered to be a finished and unchanging document.) They are tangible items fixed in time. It is paradoxical that these items are the products of people and organizations, because both people and organizations are dynamic. An inherent tension exists between the static and the dynamic.[1] Since the Internet moves beyond the characteristics of an organization or other bureaucratic structure, this tension is less obvious. The Internet takes on some of the properties of a living organism: each day documents, records, and nodes (e.g., new locations or file servers) are added to the Internet. Existing digital materials located on the Internet may be easily revised, moved to another location, subjected to text and audio annotations by people throughout the world, and evolve from prototypes to fully-functional documents. Any series of time slices would demonstrate that the Internet has evolved beyond where it was a moment before—new nodes or cells have grown and the contents within a cell have changed. In other words, rather than viewing the Internet in the more traditional sense of a network structure electronically connecting nodes throughout the world, it may be more realistic to shift this view, and define the Internet as an organism, a collection of living cells, as illustrated in Figure 1. The implication of this new view represents the first challenge for archivists: how to identify and capture a record on the Internet (or World Wide Web) in both the more traditional sense of an organizational transaction and the wider societal sense of human cultural/social interaction.

Therefore, when viewing the Internet as an organism, it is important to understand what the Internet does over and above being a "worldwide network of computer networks."[2] If the Internet is thought of as a collection of cells, it follows that the Internet is likely to mean many different things to many different users. Organizations and individuals may use the Internet for a wide variety of tasks. While some may use the Internet entirely as an information browsing tool or source of entertainment, others may use Internet materials to serve several other functions.

Hypermedia World Wide Web documents are a particularly important emerging form on the Internet because, in many respects, they illustrate the documentary potential of such media. The Internet, developed in the late 1960's, supports a wide variety of communication activities including electronic mail, bulletin boards, gophers, listservs, and commercial services (e.g., America Online and CompuServ).[3] Netscape and NCSA Mosaic are Internet-based browsing software tools that promote the exploration, display and creation of World Wide

Web materials. Both applications allow authors to link text or graphics in a document to any other document available on the Internet. Documents may contain text, images, graphics, audio resources, and digital movies, all in standard formats.[4] These applications have become popular worldwide, and can be described broadly as communications tools. It is clear that the variety of organizations creating hypermedia World Wide Web documents, the purposes of the documents, the ease with which documents may be authored and modified, and the hypertext and multimedia nature of the documents have implications for archivists and other professionals interested in documentation of the human record.[5] For example, the School of Information and Library Studies (SILS) at the University of Michigan uses hypermedia World Wide Web documents, or HomePages, to share personal information: as distance learning teaching tools with faculty and students at other institutions, as collaboration tools for high-level executive meetings, as the primary source of newspapers or magazines, as image databases, as digital libraries, and as an experimental platform for collaborative work.[6]

The most commonly found types of World Wide Web documents might be characterized as personal and project HomePages. Personal HomePages have become a popular venue for sharing a wide variety of personal information with the entire Internet community. Many individuals include images that describe some aspect about themselves, a narrative description of their interests or cur-

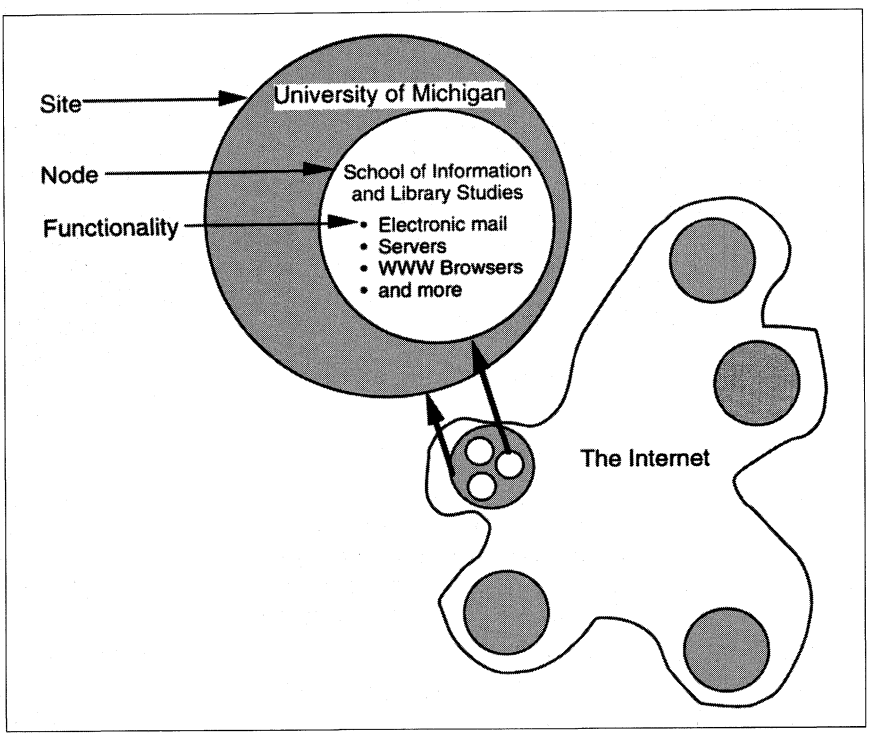

Figure 1
Conceptual Representation of the Organic Nature of the Internet

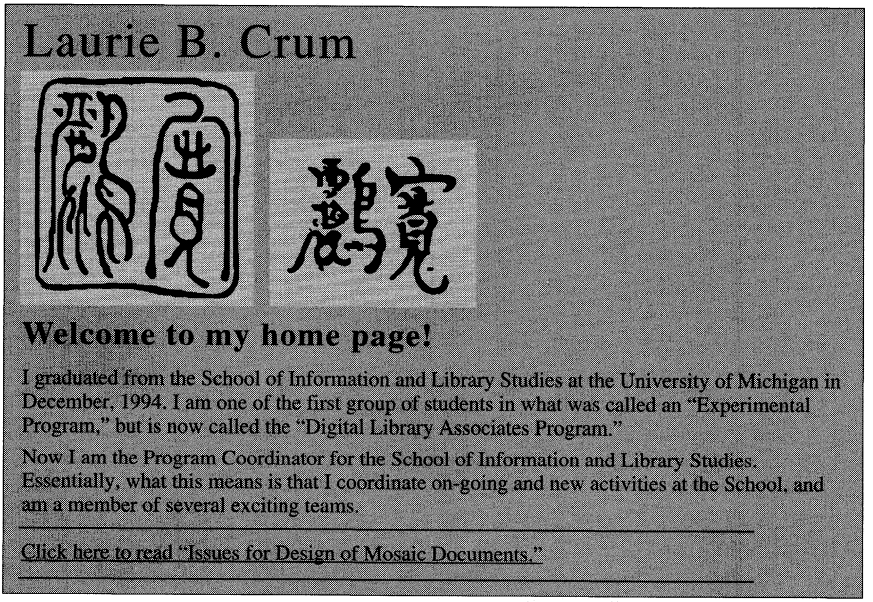

Figure 2
Example of a Personal HomePage

rent work, and hyperlinks to other documents such as papers and resumes (see Figure 2).

The purpose of a project HomePage is to make information about a particular project available to a broad and general audience. An example of this document type, the project HomePage for the NSF/NASA/ARPA Digital Library Project being developed at The University of Michigan, is shown in Figure 3.[7] Readers are encouraged to browse through the material organized in this document by choosing hypertext links. Each link provides information about different aspects of the project, including a copy of the original proposal, the current status of the project, project participants, and working papers written by team members. The project HomePage serves as both a browsing tool for interested readers and as documentation of the project.

Hypertext Mark-up Language (HTML)

HyperText Mark-up Language, or HTML, is the tagging language currently considered to be the primary format for hypermedia World Wide Web documents. Document structure and text styles are indicated by applying a set of predefined tags to either side of a block of text. For example, an author might indicate a major heading for a section with the following tag:

<H1>Introduction to Hypermedia World Wide Web Documents</H1>

The tag "H1" indicates a major header of level "1." The tag is enclosed in brackets (< and >) and the end of the tag is indicated by a backslash (/). There are currently six levels of headers available to designers through HTML. There are also several additional text style tags which allow the designer to create

many styles such as lists, quotation blocks, and bold or italicized text. Using a browsing tool which interprets HTML, this major heading would appear to the viewer as:

Introduction to Hypermedia World Wide Web Documents

In this manner, authors may easily create and globally disseminate hypermedia World Wide Web documents which contain images, digital video, audio, and hypertext links to other documents. Part of the immense popularity of the World Wide Web may be attributed to the ease and speed with which hypermedia World Wide Web documents can be created.[8] It is important to note that HTML itself is evolving. Since the introduction of "NCSA Mosaic," the first hypermedia World Wide Web browsing application, new versions of HTML with added capabilities have been released. What might be labeled a "grass roots" or common, easy to use language is evolving to a more complex, standardized language. As the complexity and capabilities of HTML are increasing, so too is the potential of more structured, predictable document types and forms. The second challenge for archivists, therefore is to establish and maintain leadership positions in defining the purpose and use of these continually evolving documents.

Books and Hypermedia World Wide Web Documents

Creation and use of new forms of digital documents have been termed by some as a digital revolution. When accounting for the whole of human history, however, it may be more correct to frame global digital communications a part of a continuous literary evolution. One important step in this evolution, the cre-

Figure 3
Example of a Project HomePage

ation of the book, provides interesting cultural parallels and insight to the current digital evolution. The results of these observations yield even more challenges for archivists and other professionals involved with the processes of understanding and influencing new document forms.

The book, as we know it today, was three centuries in the making. Like the current digital evolution, the evolution of the book was very dependent on the technologies (materials, inks, etc.) of the time. Although modern book designers continue to face challenges similar to their ancient counterparts (e.g., layout of materials, the use and placement of illustrations), a set of specifications has emerged for books which assist book designers in constructing materials that are easy to use and read.[9] In fact, books are predictable because the evolutionary nature of the development of the book has yielded a great deal of well tested and refined specifications for their physical structure, front matter, text and type conventions.

The physical structure of a book is fairly flexible in terms of characteristics such as size and color, however, there are detailed specifications for the jacket, inside jacket, binding, and end papers. Standard mechanisms exist for the front matter that allow the book designer to display, in a predictable way, the name of the author, the publisher, copyright information, features and content of the book, the rationale behind the book, and additional works by the author. The text or body of a book also tends to be structured and organized consistently both within a book and among books making for distinct relationships among the elements of a chapter, section and subsection.[10]

When designing books, the mission of the typographer is to construct an effective graphical method of expressing both the semantic functions and the intellectual content of the work. Selection of typeface, as an element in graphic design, is an important step in the design process and a critical component in the overall effect of a book. There are five typographic variables: typeface, type size, type body (size and leading), measure (width of a full line) and paragraph indent. By integrating and combining these variables, the book designer constructs the structure of the content for the book.[11]

Like the book, the physical structure of hypermedia World Wide Web documents has flexibility. As books may vary in size, color and shape, hypermedia World Wide Web documents may be viewed from a variety of different computers, monitors, and browsing software. Using HTML, a hypermedia World Wide Web document, may contain a variety of text tags. The primary tags are title, heading and paragraph. Text may be formatted in several ways including lists, definitions or quotations. The document may display graphics, images, digital video, digital audio and annotations. Hypertext links allow a hypermedia World Wide Web document to be connected with other hypermedia World Wide Web documents.

The front matter of a hypermedia World Wide Web document is limited by a 64 character (or less) title tag. The document body may include information similar to the front matter of a book, but there are no standards for placement of these identifying characteristics. One document may locate copyright information at the end of a document; another may locate it in a hypertext link. The text structure of a hypermedia World Wide Web document is invented by the author through the use of the title, heading, and paragraph tags.[12] While there may be an internal structure to a particular document, no structural standards exist

among hypermedia World Wide Web documents. In fact, a well organized document may link to a document which has an entirely different structure. Further, these documents may even defy discernment of their structure due to non-linear links to other sections within a document or to completely different documents. The reader of a hypermedia World Wide Web document cannot anticipate or predict the organization of the document.

Currently, HTML tags are separated into two categories: logical style and physical style. Physical style refers to the attributes a designer can attach directly to text, such as bold or italics. Logical styles describe the role of text within a document, such as a citation or a definition. Hypermedia World Wide Web document designers are encouraged to use logical rather than physical styles to organize a document.[13] The limited set of tools available to hypermedia World Wide Web document designers is constrained by the fact that ultimate control over the font and type choices remain in the hands of the user. Unlike the book, which is a fixed medium, a designer may select specific fonts, sizes, and colors for a hypermedia World Wide Web document which may be altered by the user.

It is clear that books and hypermedia World Wide Web documents exhibit fundamental differences in terms of the stage of evolutionary development, the time frame of each evolution, and the impact of each tool at a global level. Book design has become an extremely well defined and standardized process. Arbitrarily creating and enforcing standards at this stage may actually inhibit rather than foster the development of techniques and tools for hypermedia World Wide Web document organization and construction. Hypermedia World Wide Web documents are in the early stages of development. It is clear from examination of the evolution of the book that the modern book developed from cultures experimenting with new forms and technologies.[14] The purposes, uses, and creation of hypermedia World Wide Web documents must also evolve as technology and culture evolve. Archivists should plan for, permit and even encourage a spirit of experimentation and innovation. It is important to remember that the book took three centuries to develop. Considering the rapid changes in current technology, it is clear that it will take far less than three centuries to develop standards and an understanding of hypermedia World Wide Web documents. It would be a mistake to develop standards early in the process that may stunt or halt further growth or evolution of hypermedia World Wide Web document design and use. Further, reliance on static media, such as books, to provide specifications and methods for document creation may promote use of hypermedia World Wide Web documents as static rather than dynamic media.

Finally, the dissemination of hypermedia World Wide Web documents has a far greater immediate impact than that of the book. Books and book standards spread slowly from one culture to another. Hypermedia World Wide Web document use has rapidly increased in a very short amount of time. Communities and cultures throughout the world are creating and reading each others hypermedia World Wide Web documents. New cities, states, and countries are being added to the World Wide Web daily. Any emerging specifications or standards are likely to have an unpredictable, profound impact on a great number of cultures.

Archivists must balance the opportunity to take the lead by defining structures and specifications for hypermedia World Wide Web documents with the role of observing, shaping and guiding the use, purpose and definition of these

documents. A useful definition for hypermedia World Wide Web documents must be flexible and dynamic in order to account for cultural and technological developments. A third challenge for archivists lies in defining and preserving hypermedia World Wide Web documents of enduring value.

Defining Hypermedia World Wide Web Documents

Are hypermedia World Wide Web documents records? Like other electronic materials, they do not have a clear definition of their "recordness" in traditional archival and legal terms.[15] David Bearman has proposed a definition that suggests an electronic record is any communication that crosses a business transactional boundary between two entities (human or machine).[16] Using this definition, most materials available on the Internet become records when they are viewed by another person or the document header is scanned and selected by a World Wide Web based search engine. If a hypermedia World Wide Web document is defined by its transactional nature, it might be said that every hypermedia World Wide Web document on the Internet is a record. However, since each cell of the Internet has different uses and goals for its materials, it is entirely possible that the Internet provides much of the overall context for the definition of a document itself. The author proposes, therefore, a definition of a hypermedia World Wide Web document that incorporates not only the concept of transactions, but also the placement of the document within the wider context of the electronic and societal environment. Such a definition is not only useful in setting the context for observing and understanding emerging document forms, a necessary element for archival appraisal, but is also adaptable to other electronic communications documents.

This proposed definition of the informative value of a hypermedia World Wide Web document is represented by Figure 4. It is based on functional requirements for electronic record keeping developed by David Bearman, and concepts discussed by Richard Cox, James O'Toole, Hugh Taylor, and Donald Norman combined with this author's own observations about the organic nature and evolving use of the Internet and hypermedia World Wide Web documents in particular.[17] The core of the definition, the concept of informative value, is derived from five elements: complete, symbolic value, meaningful, maintained, and available. These elements explain, expand and build the core concept. The importance of these elements or characteristics can be more or less significant depending on the function of the material, the goals of the organization, and the mission of the appraising archivist. The most important implication for archivists attempting to manage such documents is that by using this definition, the distinguishing characteristics between a document and a record may become less noticeable, or even unimportant, depending upon what the archivist is seeking to document. A single World Wide Web HomePage may contain materials which satisfy current definitions for both records and documents. For example, a project HomePage may contain: continually updated materials which describe the activities of the project to the organization (e.g. reports to sponsoring agencies); general information regarding the project intended for a broad audience; and links to information that is completely unrelated to the project (a link to a HomePage about cats from the HomePage of a principal investigator of the project). All of these materials may have long-term value to the project. Further,

individuals from both within the project and outside of the project may contribute to the HomePage.

The specific definitions of each element are outlined below:

Document with Informative Value. To have enduring value, the document needs to provide some substantive information that is valuable to the organization or individual creating the document, or the primary or secondary audiences reading the document. In other words, the document should have a goal or purpose, and the design of the document should reflect all of these notions.[18] For example, the information available from the SILS HomePage include materials about the program; documentation of current research projects; course descriptions, assignments and syllabi; profiles of the students, faculty and staff at the school (and their roles); and Internet resources developed at SILS. These materials have clear value to SILS administrators, the students, faculty and staff, and other Internet users, although the value may be different for each group. The SILS HomePage has multiple goals which include providing a location for students to share personal information, documenting the progress of research projects, providing course information to students, displaying information about the SILS program, providing links to additional resources, and highlighting the SILS program.

Complete. Defining a document as complete can be relatively easy in the paper world. However, as Richard Cox has asserted, in the electronic environment it becomes increasingly difficult to discern when one document ends and another begins.[19] For the purpose of hypermedia World Wide Web documents, it can be said that most documents branch from the first screen of a HomePage. The document itself can contain many elements: images, audio files, digital

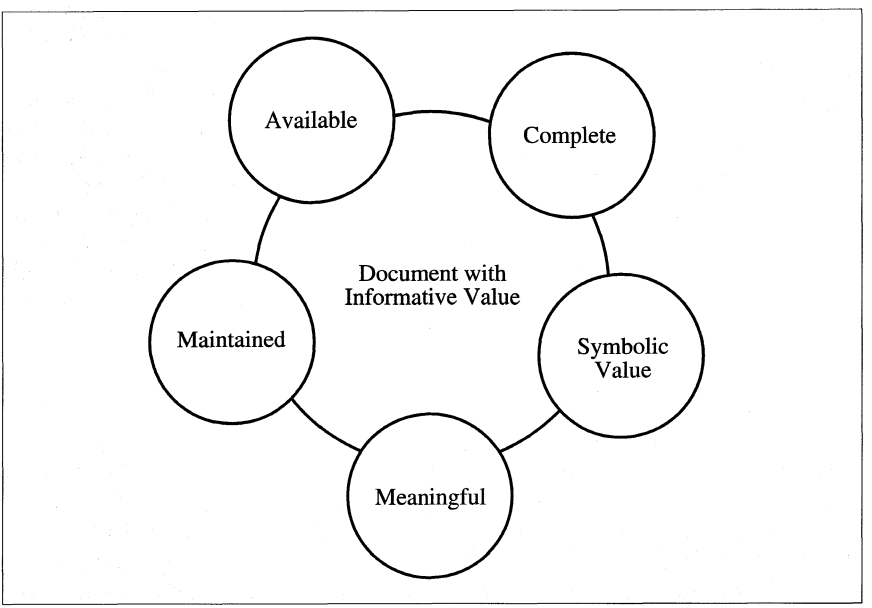

Figure 4
Symbolic Representation of a World Wide Web Document Definition

video files, and links to other documents. It is not as critical that each document has a concrete ending, as each link takes the user to another complete document. For example, when a user selects a hypertext link, that link should take them to a document rather than a blank page that states "Information will be added soon." Therefore, a hypermedia World Wide Web document has a clear beginning, but no clear ending. The critical point is that there are clear paths, and each is a valid link with informational content. This requires not only that authors monitor their own links to other documents, but also that there be a strong commitment by Internet cells to maintain these links and capabilities in an environment where even the documents themselves are continually changing.

Symbolic Value. James O'Toole has examined an important value of a record that is rarely highlighted in the electronic environment: the human motivation for making and keeping a record. Personal motivational aspects for creating and storing a record differ from organizational or governmental motivations. Such motivation could be as simple as leaving a legacy, a common psychological process, or it could be more complex, such as maintaining a sense of omniscience or power (e.g., a religious record). Ultimately, O'Toole suggests that records should be viewed for the nature of the role that they play(ed), or the symbolic value of the record.[20] Since individuals and organizations make a conscious choice to create and disseminate many electronic documents, especially hypermedia World Wide Web documents, this choice should be considered in the definition of the document. In fact, many hypermedia World Wide Web documents are highly "artifactual" when viewed in this light, that is they represent the values of a person or a group at a particular point in time.

Meaningful. Numerous reasons exist why an individual or organization might choose to create a document. In some situations, for example, a corporation, all hypermedia World Wide Web documents may be created for business purposes. At the University of Michigan, both individuals and organizations are encouraged to create hypermedia World Wide Web documents. The meaningfulness of the document should be attached to either the individual or the organization. An individual may create a document, for example, the beer page, that is extremely meaningful to him/her, but not to The University of Michigan. An organization, such as SILS, may create the SILS "HomePage" that is meaningful to SILS. Whether or not these documents are meaningful to the entire Internet community is irrelevant. Therefore, the value of a hypermedia World Wide Web document is attached to the creator of the document—whether the creator be a representative of an organization or an individual (i.e., its provenance). Like current users, subsequent or future users may or may not find the documents on the Internet to be meaningful.

Maintained. Another important characteristic of an electronic record is that it is maintained. Many documents and a tremendous volume of information are available on the Internet. As the number of organizations and individuals using the Internet grow, so too will the amount of materials available. It is critical to information users that the creator maintain the information in hypermedia World Wide Web documents or retire them. Documents whose purpose is to maintain current information become useless once the information is no longer valid. An additional concern is the lack of authority control over documents available on the Internet. Since no global formal method exists for regulating or

validating document integrity in terms of authorship, individuals and organizations should strive to provide internal mechanisms for ensuring that proper controls are placed on documents. A second reason to promote organizational knowledge of the authors is that the organization is the only "cell" which will know whether or not a person who created a document is still affiliated with that unit. It is easy to imagine an Internet in 20 years that has millions of documents created by individuals who are no longer alive, or who are located in different organizations.

Available. Availability over time in this instance is not simply a matter of long-term preservation. It is important to note that hypermedia World Wide Web documents may be available to anyone who has access to the Internet, or they may be restricted via the use of software (password or client machine address) or hardware (no Internet connectivity). The overall definition of a hypermedia World Wide Web document should consider whether the intended primary audience is the entire world, or a known organization. Although the Internet has traditionally been a forum for "free" access to information, developments of commercial uses of the Internet have already begun and will likely continue as the Internet and the information superhighway grow. Despite the inevitable commercialization of the Internet, however, it may still be the last frontier for the "candid" document. The Internet has enjoyed a history of nonregulation, and is a forum for any individual who has the ability to connect and access information. Documents across the world reflect the daily life and concerns of schools, businesses, institutions, and individuals. From this view, hypermedia World Wide Web documents and other documents on the Internet, may prove an invaluable form of social and cultural documentation to archivists and historians.

While the definition outlined here may be valid for the moment and provide a model for appraisal of electronic documents, it is likely to change and grow as the Internet develops. It is difficult to imagine what the Internet and its documents, and computing environments will be in 20 years. If the critical issues described above are not addressed by concerned communities such as archivists, however, it is almost certain that the future Internet will no longer be a useful tool, but rather another global mess.

Conclusion

It is clear that global networking will alter the way people and organizations communicate; create, share, find, and maintain information; and indeed the way they live. Archivists who deal with the records and documentation of organizations and society need to grasp the significance of and participate in these tremendous changes. Users of the Internet and World Wide Web have seized the opportunity to create and disseminate vast amounts of information. Archivists are in a unique position to shape the development of World Wide Web and help foster its growth as a useful tool. To do this it is necessary to observe and develop an understanding of global, flexible, dynamic, virtual organizations; evolving document technologies and forms; and new digital document types. The worldwide implications of the digital evolution are astounding. Archivists and other professionals must be careful to guide the growth of this new form of communication (and literacy) without imposing arbitrary or ill-

timed standards and specifications. It is time for archivists to begin to parent the digital evolution.

ABOUT THE AUTHOR: Laurie Crum completed her Master's degree as a pilot student in a newly developed Digital Library Associates Program at the School of Information and Library Studies of the University of Michigan in December, 1994. She is currently the coordinator for current and future activities, programs, and sponsored funding at the School of Information and Library Studies at the University of Michigan.

NOTES

1. John Perry Barlow, "The Economy of Ideas: A Framework for Rethinking Patents and Copyrights in the Digital Age (everything you know about intellectual property is wrong)." *Wired* (March 1994): 85-90 and 126-129. The author broadened the notion that copyright of electronic information should be dynamic rather than static to include even the documents themselves.
2. Michael Fraase, *The MAC Internet Tour Guide: Cruising the Internet the Easy Way* (North Carolina: Ventana Press, 1993). There are many sources available regarding the Internet. The notion that the Internet is a network rather than an organism is common to most definitions.
3. Cindy Morrow, ed., *The Internet Unleashed* (Indianapolis, IN: Sams Publishing, 1994).
4. The Beginner's Guide to HTML, often called the classic work on the subject is located at the following URL:
   http://www.ncsa.uiuc.edu/General/Internet/WWW/HTMLPrimer.html.
5. The diversity of organizations and activities being developed on the World Wide Web may be surmised by a quick glance at the following World Wide Web Pages:
   What's New?: http://home.mcom.com/home/whats-new.html
   What's Cool?: http://home.mcom.com/home/whats-cool.html
6. The SILS HomePage is located at the following URL:
   http://www.sils.umich.edu/.
7. The NSF/ARPA/NASA UMDL HomePage is located at the following URL:
   http://www.sils.umich.edu/UMDL/HomePage.html
8. Gary Wolf, The (Second Phase of the) Revolution has Begun, *Wired* (October 1994): 116-156.
9. Adrian Wilson, *The Design of Books* (San Francisco, CA: Chronicle Books, 1993).
10. Jeff Barry, The HyperText Markup Language (HTML) and the World-Wide Web: Raising ASCII Text to a New Level of Usability, *The Public-Access Computer Systems Review* 5:5 (1994): 5-62.
11. Richard Southall "First Principles of Typographic Design for Document Production," *TUGboat* Vol. 5:2 (1984): 79-90.
12. Stanley Rice, *Book Design: Text Format Models* (New York: R.R. Bowker Company, 1978).
13. Barry, 1994.
14. Leila Avrin, *Scribes, Script and Books* (Chicago, IL: American Library Association, 1991).
15. From 36 *Congressional Federal Register*, 1234.1.
16. David A. Bearman, "Record-keeping Systems," *Archivaria* 36 (Autumn 1993): 16-36.
17. Ibid.

    Richard J. Cox, *Archivists, Electronic Records, and the Modern Information Age: Re-examining Archival Institutions and Education in the US with Special Attention to State Archives and State Archivists*, Ph.D. Dissertation (University of Pittsburgh, University Microfilms Incorporated, Michigan, 1992).

    James O'Toole, "The Symbolic Significance of Archives." *American Archivist* 56 (Spring 1993): 235-255.

    Hugh Taylor, "My Very Act and Deed: Reflections on the Role of Textual Records in the Conduct of Affairs," *American Archivist* 51 (Fall 1988): 456-469.

    Donald A Norman, *The Design of Everyday Things* (New York: Doubleday Currency, 1988).

18. Laurie B. Crum, "Issues in Design of Mosaic Documents." School of Information and Library Studies, University of Michigan. June 1994. Available at the following URL: http://http2.sils.umich.edu/~lcrum/design.html.
19. Cox, "Archivists, Electronic Records."
20. O'Toole, "The Symbolic Significance of Archives."