

Accessibility and Integrity of Networked Information Collections

July 5, 1993

Clifford A. Lynch

This is an Office of Technology Assessment outside contractor report. It does not necessarily reflect the views or findings of the Office of Technology Assessment or the Technology Assessment Board.

1. Introduction

Objectives and Purpose of This Paper

We are entering an age when a great deal of information is available in electronic formats and can be obtained through computer communications networks. Sometimes, such collections of network-accessible electronic information are referred to as "digital libraries" (although, as discussed later, I view this terminology as somewhat misleading; indeed, one of the issues explored here is the developing roles of such electronic information collections and their relationships to institutions such as libraries). These new electronic information resources are still being defined and almost everything about them is poorly understood; conventions, expectations, social customs, institutional roles in society, business relationships, legal precedents and public policy choices based on print and mass media broadcast environments may no longer be directly applicable in the networked information environment. This new environment challenges us to reconsider assumptions and to rethink many aspects of the current systems of information creation, management, organization, preservation and use, as well as the broader public policy context of these systems.

This paper, prepared under contract to the United States Congress Office of Technology Assessment, is intended to consider questions related to the integrity and accessibility of these new electronic information resources. These are key areas that pervade many of our assumptions about all forms of information; the resolution of integrity and access questions in the electronic information environment will play a major role in defining how we create, use and manage these new forms of information and the economic system that will support these activities, the purposes it will serve within our society, and the functions that various institutions will need to carry out in this new environment. Both integrity and access are complex issues, and have intricate

relationships not only to each other but to legal and public policy considerations such as privacy and intellectual property. Further, integrity and access cannot be considered outside of the context of the communities and institutions that create, distribute, organize, manage and use information; this leads to a discussion of the evolving roles of libraries, publishers, and the authors and readers of various types of material. Finally, because the growing flood of available information is increasingly overwhelming and unmanageable without mechanisms and conventions to allow us to identify, select, and cite particular works, I have devoted considerable space to exploring how these mechanisms are developing in the world of electronic information resources.

While there is some emphasis here on printed information and its electronic descendants, and particularly scholarly communication and publication, I have tried to at least begin an examination of the broader issues related to mass communications, newspapers and news broadcasts, government information, and new multimedia information resources that might be of interest to the general public rather than just the research and higher education communities. The reader should recognize that my comments in these areas are at times rather speculative; the research and higher education communities have been early adopters of networked information, and while even within those communities the issues are far from resolved, we have much more experience with scholarly information than we do with mass market network-accessible information.

To a great extent, this paper presents a personal view of the developing issues. My background is large research libraries and the higher education community. Other stakeholders in the areas discussed here, such as publishers or public librarians, may well have other views, and certainly even within my own communities there are many different opinions not only on what should be done but even the significance of many of the trends and developments discussed here. While I have attempted in various places in this paper to at least indicate the existence of other viewpoints and sometimes to sketch the rationale behind them, the reader should not view this paper as a fully balanced survey and synthesis of the various viewpoints on the issues. Finally, I must stress that I am not an attorney, and thus my comments on legal issues should be viewed as probably better informed about the implications of current legislation and case law on systems of information creation, dissemination, preservation and use than on the legal specifics themselves. While this paper has benefited greatly from reviewers more learned in the law, I may well have misunderstood or overlooked specific legal issues despite their efforts to help.

I have not provided an explicit executive summary of this paper. Those interested in obtaining a quick overview of the paper's coverage and conclusions should read the subsection directly following this one which gives an overview of the paper, and then proceed to Sections 15 and 16, the conclusions and recommendations for possible follow-on actions. These proposals are somewhat limited in nature; my purpose in this paper is primarily to illuminate the relationships among technological developments, electronic information, legislation and public policy, and various institutions such as libraries, and to identify and define the issues that new technologies are bringing to the fore. In a number of areas I have suggested consideration of a review and reexamination of current public policies, legislative positions, and government and private sector investment decisions in light of the trends described here; but in cases of public policy and legislation I have focused on providing information that could help to

inform such a review and reexamination rather than attempting to suggest specific outcomes. I have mentioned a few specific areas where I felt that there was a need for funding, to move the implementation of existing policy directions.

Overview of the Paper

The paper begins with a survey of recent developments in networked information resources and tools to identify, navigate, and use such resources (Section 2). This is a vital context and source of examples for the remainder of the report. As these new information resources are surveyed, this section also examines the idea of “digital libraries” and the relationship between today’s libraries as organizations and collections of electronic information. Those readers unfamiliar with the rather extensive and quickly evolving developments in networked information resources may find this section of the paper particularly difficult reading, heavy with acronyms and references to a bewildering array of organizations and systems. To help those readers, I have provided a brief glossary of terms that I felt might be unfamiliar and a short annotated bibliography of useful background readings (as distinct from specific references cited in the text).

Building on the context established in Section 2, Section 3 provides an overview of the issues involved in access and integrity questions and the relationships among them. The boundaries of the paper are also largely defined here: for example, issues of network access (as opposed to access to information on the network) are excluded, as are most of the public policy issues specific to government mandates to provide access to government information. If access is defined somewhat narrowly, integrity is treated quite broadly and encompasses not only conventions and audit trails needed to ensure consistency and accountability in the scholarly, historical and cultural records, but also questions of content bias and accuracy. The links between access and integrity are stressed: for example, ensuring the integrity of a corpus of publication is meaningless if there is no access to this body of information.

The paper then discusses the changing legal framework that governs use of electronic information as contract law rather than simple sale within the context of copyright law becomes the dominant model for acquiring access to electronic information (Section 4). This shift is shown to have profound implications for access to information and also for the integrity of the historical, scholarly and cultural record that “published” information has traditionally represented. The effects of this shift on libraries and the interlibrary loan system, which has historically been a key part of the library community’s strategy for providing their patrons with very broad access to information, is examined in depth. This is followed by an exploration of the new and even more complex questions raised by image and multimedia content in electronic networks; here we start from an environment of much greater ambiguity when applying copyright law, and find both libraries and rightsholders facing substantial problems in understanding how to manage these materials. Because of this, we find that the shift to contract **law** offers stronger operational benefits for those institutions (including libraries) that want to acquire access to image and multimedia information, although this shift again raises serious public policy issues.

The role of secondary information sources in providing access to, and evaluation of, electronic information is examined from several perspectives in Section 5; these include the role of the extensive and well-developed marketplace that exists today in databases (for example, so-called abstracting and indexing databases) and other tools that provide access to the print literature (and in future to electronic source material) and the potential impact of new tools derived from technologies such as automatic indexing. Appropriate application scenarios for the different access tools are discussed, as is the growing power of these access tools in defining the part of the “published literature” that will be read and evaluated by various communities. This power places great responsibility for quality and integrity on the secondary information providers and access tools and thus plays a significant role in establishing the overall accessibility and integrity of the published literature. Considerable attention is given to the implications and impact of quality and integrity problems in this area.

The paper then turns to the central issues of the historical and scholarly record and the access and integrity questions that surround it as much of the information that comprises this record shifts to electronic forms (Section 6). Much of the theme here revolves around the multiple uses to which this record is put, the different expectations and requirements of the various communities that rely on this record, and the social, business, legal and technical issues involved in trying to address these somewhat conflicting needs. The implications of the shift from sale under copyright to contract law based licensing emerge clearly here as a potentially serious threat to both access and integrity for this record, and help to frame the public policy issues involved in ensuring its preservation.

A series of technology developments, economic factors and market demands have led to the creation of ever more specialized publications; in the print world this is often termed “micropublishing” while in the broadcast industries it is sometimes referred to as “narrowcasting.” The trend towards increasingly personalized information products, both in print and digital forms, combined with the appearance of new electronic information sources such as news feeds or sensor data streams that can essentially be “personalized” by the recipient (or the recipient’s organization) using digital technology again raise serious challenges to the accessibility and integrity of the electronic record. They create enormous problems for libraries as the traditional managers of this record. And they raise new, complex public policy problems related to equality of access to information sources, and the implications of inequitable access by various sectors of the population. Section 7 explores this area.

The relationships between privacy and access in the electronic environment are very complex. Do users have a right to anonymous access to information resources, or at least a reasonable expectation that records of their information seeking and use should be kept confidential? Complicating these questions are conflicting requirements for accounting, cost recovery, system management and tuning, ensuring the security and integrity of the underlying computer networks, and the economic motivations of various parties to collect usage and demographic data for potential resale or reuse that run counter to the long standing policies of confidentiality and anonymity that have been established within the library community. At the same time, technological developments are in some cases preempting (at least in the near term) the as yet unresolved policy debates. Section 8, on privacy and access attempts to summarize the issues in these areas and to reflect some of the competing pressures.

Much of the information currently available through computer networks is “free”; that is, the user is not charged for using it. Section 9 examines some of the implications of free information, such as expectations about accuracy, timeliness, and value. In a real sense, the extensive reliance upon free information sources is shown to add a certain instability to the networked information environment because of the ease with which inaccurate information can quickly spread with little accountability. While the public policy questions here seem to be rather limited, this is important to developing a full picture of the networked information environment.

The paper then considers the nature of electronic works and how access to such works can be provided in Section 10. Here, one major theme is the current tendency to intertwine content and access mechanism. This has serious implications for long-term access and preservation of these works as the information technology environment continues to change. There are also subtle integrity issues that arise as we attempt to define the distinctions between content and presentation of that content. This section also emphasizes the importance of establishing and encouraging the widespread adoption of appropriate standards that allow information providers and users to separate content, presentation, and access or viewing tools for electronic works.

While we realize intellectually that photographs can be altered, the visual evidence provided by photography has provided a very important part of the historical record in our society. We have made very strong intuitive assumptions about the integrity of photography. The section on digital imaging and the integrity of the visual record (Section 11) summarizes how the development of new digital imaging technologies calls this role into question and places much greater emphasis not on the image as artifact, but on the use of verified true versions of images combined with formal, auditable links from that image to the individual or organization that has created it. This serves as motivation for the Section 12, which deals with the authentication or verification of electronic works and their creators. This requirement is a cornerstone of the integrity of electronic information; while perhaps most visible in the context of images, it pervades the use of all types of electronic information. A discussion of the issues involved in making such authentication possible leads directly to a series of issues concerning cryptographic technology, including standards, intellectual property rights and export controls (which in turn are related to the search for an appropriate balance between national security concerns and the needs for privacy and authentication in the networked environment).

The final sections of the paper consider two related issues that are again central to both access and integrity in the electronic information environment. The first, covered in Section 13, has to do with identifying and citing electronic works, and summarizes requirements, architectural approaches and standards developments to address these needs. The second issue is the intellectual identification of networked information resources; here the development of catalogs and directories for these resources is considered, with some emphasis on the role of libraries on the one hand in creating these directories and catalogs and, conversely, the way in which such directories and catalogs will integrate with existing tools used by libraries to provide access to the print literature during the long transitional period where both electronic and print information are essential parts of the scholarly communication system and co-exist. Section 14 addresses these questions.

The paper divides its conclusions into two parts. The first concluding section (Section 15) tries to summarize and tie together the various developments, trends and issues that have been surveyed, and also to set the conclusions of this paper in a broader context. One point that is emphasized in this section is that two of the key areas slowing progress in infrastructure development for networked information—standards development and the deployment of integrity and authentication support services based on cryptographic technologies—call for issues well outside the primary scope of this report to be addressed, but until these issues are addressed, will continue to cause problems in our ability to resolve access and integrity questions related to networked information. Section 16 builds on and Section 15 and enumerates specific issues and projects where actions—by Congress, by various government agencies, or by various groups within the stakeholder community—should, in my view, be considered.

Acknowledgments and Disclaimers

This paper has benefited greatly from the careful work of a large number of reviewers some known to me and others anonymous. Their comments greatly improved the quality and accuracy of the work. I thank them for their efforts and their thoughts, while taking all responsibility for errors and for opinions with which they may well disagree. I owe particular thanks to Joan Winston of the Office of Technology Assessment for multiple reasons: for inviting me to explore these extremely interesting issues in the first place; for a very careful and thoughtful review of a draft of the paper; and for her patience as deadlines slipped and the paper continued to grow and evolve at its own pace towards completion. Conversations with any number of colleagues, both in person and electronically, have been of immeasurable help in understanding these issues. Particular thanks are due Steve Cisler, Michael Buckland and Howard Besser. Various electronic newsgroups and discussion lists have also been very valuable to me in preparing this paper, both in terms of discussions (where typically I listened much more than I contributed) and by providing direct object lessons about several of the topics discussed in the paper. I would also like to thank Cecilia Preston both for a careful review of a draft of this paper and for her help with the citations and references.

References and citations for a paper of this type are a huge problem. I have not even tried to be comprehensive in the citations; while a comprehensive bibliography of developments in networked information would be extremely useful, compiling it would be a truly enormous effort, and, given the rate of developments in this area, it would need continual updating. Worse, a number of the projects have not yet been much described in the published literature, and what material exists is scattered across obscure conference and workshop proceedings and technical reports. Thus, I have had to satisfy myself with the inclusion of a limited number of references suggestive of the type of work going on today, particularly in Section 2 on the networked information context, and I would beg the indulgence of those authors who have made important contributions to the literature in this area but do not find their efforts reflected in the citations here. Some readers may find the bibliography in [Lynch & Preston, 1990] a helpful supplementary source. In sections this paper touches on other large and complex areas, such as the technology of cryptography; for basic citations in this area (as well as a very readable overview of some of the technology) the reader should see [U.S. Congress Office of Technology Assessment, 1987]. Similarly, I have not provided