Book Reviews

Arms, William Y. *Digital Libraries*. Cambridge, Mass.: MIT Pr., 2000. 287p. $45, alk. paper (ISBN 0262011808). LC 99-14773.

This important, but difficult-to-read, book is likely to elicit a broad range of responses and reactions from its readers. A tradition-minded librarian might easily be tempted to think at once about early retirement, whereas someone eager to explore the ever-changing panorama of new technologies will excitedly welcome its vision of the future.

The first chapter, a broad gauge tutorial, probably designed mainly for academic administrators or university governing boards, is unlikely to tell active librarians anything they do not already know. The following three chapters describe the development of the Internet, detail with considerable accuracy the sometimes-contentious relationships between libraries and publishers, and discuss contemporary innovation and research. Arms correctly points out that many people who grew up with the Internet believe they have "discovered" online information systems and are completely unaware that the concepts forming the basis of digital libraries were long ago researched and developed by the library community.

Librarians are not likely to receive well the author's statement in chapter four—"Computer scientists take research seriously"—with its clear, gratuitous presumption that librarians do not. Computer scientists, writes Arms, are the hares of digital libraries, whereas large libraries—and by extension, one can infer, their librarians—are the tortoises. Yet, there is little discussion about *why* librarianship was for so many years averse to research, an attitude constantly criticized by early pioneers of library research such as Pierce Butler and Jesse Shera. There is no discussion of the obvious arrogance among some computer scientists who, for years, held back the development of practical, user-friendly information systems. Missing from the work is any substantial consideration of the research role of the former Council on Library Resources (CLR); its work is alluded to only in passing. Arms suggests that research in information science is "at least 30 years old," but it is actually almost twice that age, dating back to the 1940s when Ralph Shaw developed and demonstrated his RapidSelector. Arms does not mention the pioneering theoretical research done at MIT in the mid-1950s by Calvin Mooers, perhaps the first person to connect a typewriter terminal to a mainframe, nor does he discuss the important developmental work on automatic indexing done over an entire lifetime by Gerard Salton at Harvard and Cornell.

Throughout the work, Arms discusses in substantial detail a variety of important, major issues. The interoperability of various computerized information systems, the costs of standardization, and the need to maintain legacy systems are given valuable and thoroughgoing discussion. Chapter five, devoted to personnel, users, and organizations, is one of the best treatments anywhere of how challenging it is to weld together disparate interests into a fundamentally restructured, unified service mechanism. The author properly rejects as naive the idea that technology "eliminates" the need for professional management of information. Quite the opposite, Arms maintains that technology more than ever demands managerial talent strong enough to administer personnel, equipment, and services that vary widely in every dimension. Librarians may not want to hear about the university library's declining importance relative to other information resources, but Arms does not shrink from stating this

fact. Citing an anonymous University of California professor, Arms acknowledges baldly that for many students, the Internet *is* the library. Chapter six on economic and legal issues and chapter seven on access management and security are models of clarity and detail. Chapter twelve, which deals with object models, identifiers, and structural metadata, could profitably have come much earlier in the book, for it lays an excellent foundation for the technical details underlying the concept of the digital library. The permanence and durability, or more accurately, the impermanence and evanescence, of digital data are topics that Arms discusses at length throughout the book. Just ten pages from the end may be the work's most important point about digital libraries and permanence, a quotation from Michael Wettengel, a German archivist: "Computer technology is made for information processing, not for long-term storage."

There are problems with this book that make it difficult to read. Overall, a choppy style lends an aura of incoherence to parts of the work; sometimes paragraphs contain simple declarative sentences so short that the reader is left wondering about the connections between them. Similarly, some chapters have weak transitions and do not segue well into each other. It is disappointing to find grammatical and spelling errors that a competent proofreader should have caught immediately. Did author and editor rely on the notoriously undependable grammar checkers that come with word processors?

A maddening feature of the book stems from its graphic design. Throughout the work are dozens of sidebar-like "panels," each containing a succinct summary of some extraordinarily important or essential technical matter (e.g., Java, the Dublin Core, Real Audio). Unfortunately, the designers chose to print these panels onto an overdense, half-tone background. Consequently, the panels are virtually impossible to read with any degree of speed or comfort. Furthermore, the arrangement almost guarantees lack of easy comprehension. Only a dedicated (or perhaps

masochistic) researcher will get through the panels, without which the full import of the author's work cannot be appreciated. It is astonishing that so distinguished a publisher would permit this intrusion into the communication process and quite ironic that it occurs in a book whose chief subject is communication over the ages. If the work is reprinted or revised, correction of its editorial and production flaws is mandatory.

In a bizarre and embarrassing error in panel 9.1, which deals with the *Oxford English Dictionary,* Arms credits its creation to "James Morris and his colleagues." It is universally known that Sir James Murray (1837–1915) almost single-handedly created the *OED*, having spent the last thirty-six years of his life on the project. (The *OED*'s home page lists all editorial personnel connected with the project from its inception; there is no reference to James Morris.)

A generally excellent glossary helps the reader understand numerous technical terms and acronyms, although some acronyms are left undecoded and entered into neither index nor glossary. For example, CNRI (Corporation for National Research Initiatives) first appears on page 90, again on page 102, but is not explained until one reaches panel 12.8 on page 240. The index itself is seriously deficient, lacking entries for obvious topics such as the CLIR (or its predecessor, the CLR), Library of Congress, NISO, RLG, RLIN, and UNIX, and excluding a host of contemporary personal names important to the development of digital libraries (e.g., Henriette D. Avram). In fact, the entire index contains only two personal names: Vannevar Bush and J.C.R. Licklider. Quite a number of index entries even lack their full complement of locators.

In sum, *Digital Libraries* is a mixed bag, marred by some serious errors and questionable claims, but serving as a powerful stimulus to thinking that departs from tradition. It integrates into a single work detailed information on a vast range of modern technologies and inventories a cluster of unpleasant problems besetting

contemporary librarianship. *Digital Libraries* is a disturbing, distressing book, as it should be. Modern librarianship can benefit from a gadfly, and Arms plays that role with admirable effectiveness.—*Allen B. Veaner, Tucson, Arizona*

Authenticity in a Digital Environment. Washington, D.C.: Council on Library and Information Resources, 2000. 76pp. $20, alk. paper (ISBN 1-887334-77-7).

On January 24, 2000, the Council on Library and Information Resources (CLIR) convened a group of recognized experts to ponder the questions: What is an authentic digital object? How do the standards of archival preservation apply to digital artifacts? Where can archivists position themselves, as the keepers of the intellectual record, to ensure that future generations will be able to study certifiably genuine digital documents? In an effort to get the discussion started, CLIR asked five leaders in the fields of archiving and digitization to write position papers on various aspects of the topic of ensuring authenticity of the digital record. This publication contains those five essays.

Charles T. Cullen, president and librarian of the Newberry Library, writes in his essay, "Authentication of Digital Objects: Lessons from a Historian's Research," of the difficulty of affirming the provenance of paper objects, let alone that of digital objects. He expresses chagrin at the lack of real signature markings that would prove beyond doubt that a document is truly the work of the assumed author. He touches on the ease with which changes can be made without note in digital work, even when an honest transcription is attempted. Who can vouch for the fidelity of the transcript to the original? Finally, Cullen urges librarians, publishers, and authors to push forward to find methods to mark digital works with identifiers that will prove authorship.

Peter B. Hirtle, codirector of the Cornell Institute for Digital Collections, in his essay, "Archival Authenticity in a Digital Age," focuses on the records used for authentication of an object. As an example, he uses the USS *Constellation,* a wooden-hulled navy vessel moored in Baltimore Harbor. The *Constellation* was recorded as being Baltimore built in 1797, and all historic documents regarding her construction and launch were consistent in this regard. However, it was discovered that she was actually built in Norfolk, Virginia, in 1854, and constructed to look like the much-admired USS *Constitution,* but commissioned for duty during the Civil War, not the Revolutionary War. The documentation had been "adjusted" in 1909 to reflect the earlier building date, although records prior to 1909 showed her accurate construction date. Hirtle's point is that if paper records can be so manipulated and still appear authentic, archivists will need to exert great effort to maintain accurate provenance for digital objects. He proposes that one collective method will be "social mechanisms of control," or diplomatics, a body of practices that has long been accepted as a rigorous validation technique. Trust in the repository will continue to be a key concept.

David M. Levy, a consultant on documents, digital libraries, and publishing, uses a humorous example in his essay, "Where's Waldo? Reflections on Copies and Authenticity in a Digital Environment," to make the point that in a digital world, in which a copy can be identical to the original (or nearly so), the ascertaining of authenticity of the original becomes almost impossible. First establishing that it may not be possible to create stable digital objects, he suggests that audit trails may be useful in defining the "original." He posits that our first step is to understand what it is we want to accomplish, then to discern what is possible to accomplish.

Clifford A. Lynch, executive director of the Coalition for Networked Information, in his essay, "Authenticity and Integrity in the Digital Environment: An Exploratory Analysis of the Central Role of Trust," sets forth many propositions