ARCHIVES AND MUSEUM INFORMATICS

Cultural Heritage Informatics Quarterly

VOLUME 10 • NUMBER 4 • 1996

Archives and Museum Informatics: Cultural Heritage Informatics Quarterly (ISSN 1042-1467) is published by Archives & Museum Informatics, 5501 Walnut Street, Suite 203, Pittsburgh, PA 15232-2311 USA; tel 412-683-9775; fax 412-683-7366.

The journal is edited by David Bearman, whose authorship may be presumed for all items not otherwise attributed.

Archives and Museum Informatics carries news, opinions, and reports on information technologies, techniques and theories relevant to archives and museums.

Submissions of press releases, publications and software for review, articles, and letters to the editor are welcome. Submissions and requests for publication guidelines should be directed to Jennifer Trant, Managing Editor, at the editorial office: Archives and Museum Informatics, 5501 Walnut Street, Suite 203, Pittsburgh, PA 15232-2311 USA; tel 412-683-9775; fax 412-683-7366; e-mail: jtrant@archimuse.com.

Beginning with Volume 11, Issue 1, 1997, *Archives and Museum Informatics: Cultural Heritage Informatics Quarterly* (ISSN 1042-1467) will be published by Kluwer Academic Publishers, Order Department, P.O. Box 322, 3300 AH Dordrecht, The Netherlands; tel +31-78-6392392; fax +31-78-6546474; e-mail: services@wkap.nl. Subscription requests should be directed to Kluwer; questions concerning submissions and manuscripts should continue to be directed to the editorial office in Pittsburgh.

Archives and Museum Informatics also publishes occasional technical reports available for purchase. Current titles can be obtained from the editorial office as well as from the Archives and Museum Informatics web site. <http://www.archimuse.com>

Archives and Museum Informatics
Cultural Heritage Informatics Quarterly

Volume 10 • Number 4 • 1996

CONTENTS

EDITORIAL

"And the last shall be first..." *by J. Trant*/311

ARTICLE

Electronic Records Research: What Have Archivists Learned From the Mistakes of the Past? *by M. Hedstrom*/313

Increasing the Acceptance of Functional Requirements for Electronic Evidence *by W. Duff*/326

Harnessing the Potential: Effective Access to Cultural Heritage Information *by L. Elliott Sherwood*/352

"Spiders across the stars"—A Web-based Model for Providing Access to Multi-Institutional Museum Information *by D. Stam*/374

"And the last shall be first..."

Jennifer Trant

This is the last issue of *Archives and Museum Informatics* to be published from the Pittsburgh office of Archives & Museum Informatics. I felt that David Bearman's decision to cease production of the journal should not go by unmarked, and offered to guest edit a 'final issue'. David (despite his reputation for modesty) did not want a *feschrift* per se, so we agreed on a series of invited papers addressing issues and ideas that had preoccupied David during the last ten years, and that had featured regularly in his writings. I contacted a number of colleagues, all of whom agreed, enthusiastically, to write something in tribute. I'm grateful to all of them for their contributions.

In the intervening time between our agreeing to this concept and its execution, Kluwer Academic Publishers acquired *Archives and Museum Informatics* and agreed to transform it into a peer-reviewed academic journal. Rather than cease publication, the journal gained a new lease on life, as a vehicle for critical discourse. I've long felt that the cultural heritage information management community needed a scholarly journal as a focus for its development, and agreed to serve as the Managing Editor of the new Kluwer publication with the hopes that I could help to create a venue for the serious exploration of issues of knowledge representation and networked access to cultural heritage information.

Rather that marking an ending, this series of invited papers, then, celebrates a new beginning. Split between this issue and Volume 11, no. 1 are a group of eleven papers that explore

questions of creating, capturing, managing and using electronic information that documents cultural heritage, whether electronic records in archives, or computerised documentation in museums and libraries. These papers explore a series of common themes, most notably the difficulty of finding information in the networked environment, the challenge of managing electronic records over time, and the need to deliver information in a useful manner. The authors look at the individual and institutional challenges that we face as we transform the way we work and do business; either explicitly or implicitly, each offers a contribution to our collective research agenda.

Not surprisingly, all of these issue hinge upon the development of adequate standards for representing the content of the information objects in our care. Metadata managers[1] face an old problem in a new guise, and with a new twist. Rather than being specialist cataloguers that develop precise representations of the knowledge of a specific discipline, we are faced with the challenge of developing descriptions that can form a part of inter-operable information architectures, enabling the integration of information across disciplines and media types. We need to think more broadly.

Any ongoing development of our practice must take place with an awareness of our actions and their implications for the future. As Margaret Hedstrom points out, we need to develop a culture of research and analysis that moves us forward in a conscious way. We're all grateful to David Bearman for his continuing contribution to our understanding of the issues we confront. Despite his withdrawal from active work on the journal, I'm sure we'll not be without his insight or opinion, and I look forward to collaborating with him in his new role of Editor-in-Chief.

[1] As Cliff Lynch observed these are the new, and hopefully more highly paid cataloguers in our midst.

ARTICLE

Electronic Records Research: What Have Archivists Learned from the Mistakes of the Past?

Margaret Hedstrom
School of Information, University of Michigan

In 1988, I wrote a brief essay for the *Archival Informatics Newsletter* called "Optical Disks: Are Archivists Repeating the Mistakes of the Past?"[1] In that article, I used the example of archivists' increasing interest in digital imaging applications within archives to raise questions about the capability of archivists to manage, preserve, and make available records created in digital formats. I suggested that the archival community should analyze some of its mistakes with electronic records in order to develop better strategies for dealing with new technologies in the future.

My 1988 article was not a reaction to the technological imperatives of optical disks. Rather, it was an assessment of where the archival community stood in relation to the archiving of electronic records at a point when yet another new technology was gaining momentum as *the solution* to recordkeeping problems. One basic lesson drawn from the experience with electronic records was that the archival community must be more vigilant with respect to emerging technologies and be prepared to influence standards, practices, and applications before new technologies achieve a degree of market penetration. Another lesson was that the short life span of storage media is not as significant a problem as software obsolescence. Archivists who

[1] *Archival Informatics Newsletter* 2:3 (1988): 52-53.

valued optical media for its longevity would still have to contend with numerous changes in storage formats and retrieval systems during its projected 30- to 100-year life expectancy. During the 1980s, as organizations introduced more complex and powerful applications, such as decision support systems, GIS, and integrated office systems, it was also becoming apparent that the standard archival practice of preserving all electronic records as flat files would not suffice as the only acceptable method for long-term preservation. Managing software dependencies had to replace software independence an archival management strategy. Finally, new types of digital objects had characteristics and capabilities not found in paper-based records. The notion of "print to paper" or keeping the hard copy for posterity had limited resonance when electronic recordkeeping systems began to diverge from the paper legacy in their creation, organization, and use.

The key problems required new approaches for moving the archival community into a more strategic position. Strategies that would enable archivists to take action when new technologies were introduced, rather than waiting for large backlogs of non-standard, unrecoverable records to accumulate, seemed like the only feasible way that the archival community could get on top of the electronic recordkeeping problem. Archivists would have to influence standards for encryption, compression, storage, and data or document exchange to avoid unacceptable costs for long-term preservation of software-dependent data. Engagement in system design to identify archival records and provide sufficient access points for them offered a practical alternative to capturing snapshots of databases or conducting customized back-end reformatting of archival electronic records.

This is an opportune time to revisit the question of what the archival community has learned from recent research and development in electronic records management and digital preservation. Digital imaging applications in archives are

maturing with the adoption of consensus standards for image quality, storage formats, and description.[2] New "virtual collections" are being planned and designed that bring together previously dispersed collections around a common provenance or theme. Archivists have also made considerable progress toward a shared conceptual framework for electronic records management based on a definition of records as transactions, analysis of business processes, compilation of the warrant for recordkeeping, and maintenance of the content, structure, and context of records.[3] Nevertheless, one cannot overlook the fact that archivists have been more enthusiastic about applying new technologies to access systems and conduct retrospective digitization within archival institutions, than they have been eager to identify, preserve, and provide access to archival records that begin their life elsewhere in digital form.

A major step was taken in the early 1990s with the recognition that systematic research on archival management of electronic records was a necessary ingredient in any long-term solution to electronic records issues. In 1991, the U.S. National Historical Publications and Records Commission (NHPRC) sponsored a conference on electronic records research issues and developed a research agenda to address systematically the problems of archival management of electronic records. The

2 For a comprehensive compilation of best practices, see Kenney, A.R. and S. Chapman, *Digital Imaging for Libraries and Archives* (Ithaca: Cornell University Department of Preservation and Conservation, 1996).

3 David Bearman summarized these developments in a recent issue of this journal. See Bearman, D., "State of Electronic Records Management Worldwide", *Archives and Museum Informatics* 10:1 (1996): 3-40. For a summary of recent electronic records projects in North America, see the reports and background materials for the conference on Electronic Records Research held at the University of Michigan, June 28 and 29, 1996. <http://www.si.umich.edu/e-recs/>

conference set a national agenda for research which posed ten priority areas for research. Research proposals were encouraged to determine functional requirements for electronic records, define standards and applications for metadata that would facilitate electronic records management and preservation, identify effective policies and program development strategies for archives and other institutions charged with responsibility for preservation and access, and reduce barriers to adoption of effective electronic records strategies in archival institutions and among archivists.[4] Three questions dealing with the functional requirements for electronic records management, the costs and conceptual issues involved in capturing and retaining data and descriptive information, and strategies for preserving software-dependent data were deemed most important as a starting point for a research program. Within the first five years, twenty-four projects were funded, ranging from the purely research-oriented project at the University of Pittsburgh to small requests for consultants to provide advice to archival institutions on starting an electronic records program.

A recent conference on Electronic Records Research, held at the University of Michigan in Ann Arbor on June 28 and 29, 1996, provided an opportunity to assess progress toward answering the questions on the 1991 research agenda. As the principal organizer of the conference, I compiled results from more than twenty NHPRC-funded electronic records projects. Seventy-two colleagues met to review recent research and revise the research agenda. The Conference Web Page <http://www.si.umich.edu/e-recs/> and the Final Report provide summaries of the projects' findings and draw general conclusions about recent research. In this article, I will share some very general observations about recent research, the limitations of its findings, and the nature of the research process itself.

Recent research on electronic records management and preservation produced models which demonstrate that it is technically feasible to create and maintain reliable records in electronic form. Although only one project, the University of Pittsburgh's project entitled "Functional Requirements for Evidence in Recordkeeping," focused exclusively on the top priority research issues, many other projects addressed some aspect of the research agenda.[5] The archival, business, and legal communities all have made important strides in recent years in defining the requirements for reliable evidence and proposing standards and models to support satisfaction of those requirements.[6] Research has also demonstrated that a variety of laws, regulations, policies, standards, and professional best practices, referred to as the warrant for recordkeeping, govern or influence the content, form, structure, and uses of electronic records.[7] For archivists and records managers, shaping electronic

4 National Historical Publications and Records Commission, *Research Issues in Electronic Records* (St. Paul: Minnesota Historical Society, 1991).

5 The most comprehensive documentation on the Pittsburgh Project is available from the project's web site. February 1996. <http://www.lis.pitt.edu/~nhprc/>

6 Within the archival community, two most developed models are the Pittsburgh Project's "Functional Requirements for Evidence in Recordkeeping" and a model being developed at the University of British Columbia called "Preservation of the Integrity of Electronic Records." <http://www.slais.ubc.ca/users/duranti>

7 The Pittsburgh Project uses the term "literary warrant" to refer to this body of laws, guidelines, and best practices. For additional details, see the Pittsburgh Project Home Page. In her Ph.D. dissertation, Wendy Duff explores this issue further and evaluates how the origins of the literary warrant affect different users' perceptions of its validity and weight. See Duff, W., *The Influence of Warrant on the Acceptance and Credibility of the Functional Requirements For Recordkeeping*, Ph.D. dissertation,

records management practices is not simply a matter of developing standards, regulations, and guidelines for electronic records management. Other authorities must be consulted and consistency established between archival requirements and other requirements for evidence and recordkeeping.

The models developed by recent research and development projects seem salient in organizations that operate under formal rules of accountability, rely on structured business processes, and carry out clearly defined transactions. Less is known about the effectiveness of these solutions in organizations that are less structured or that conduct their work with less formal processes and procedures. Recent research has emphasized preventive measures and strategies for establishing the pre-conditions for long-term preservation of electronic records, but there has been little progress toward developing solutions for preservation of electronic records when their long-term value is not recognized at the point of creation or when long-term preservation serves a perceived need that is extraneous to the business process at hand—such as for historical, genealogical, medical, or environmental research. The Pittsburgh project and Wendy Duff's recent dissertation on the warrant for recordkeeping, identified laws, regulations, and profession guidelines that encourage creation, maintenance, and use of records, but this research did not identify a clear warrant for long-term preservation of records to meet putative societal or cultural needs. Recent research has also identified risks, costs, and benefits as important criteria for deciding whether to capture a record, how and how long to maintain it, and how to make it accessible. But there is a paucity of research on risks or on the cost factors and benefits of various approaches to electronic records management and preservation. All of this suggests that while the archival

(Pittsburgh: University of Pittsburgh, 1996). *(Editor's Note: A section of this research is reported in this issue.)*

community has generated some substantive answers to significant research questions, we have far from exhausted the potential areas for research.

Putting the results of research into practice will entail significant changes in the practices of archival institutions and individual archivists. The research on the warrant for recordkeeping demonstrates that archivists and records managers have to engage a broader set of authorities to justify the ends of long-term preservation and suggests that archivists acting alone cannot bring about needed changes in policy, system design, management, and practice. Recent experience also confirms that electronic records management and preservation issues are best handled when they are considered during the design and development of new systems, but that most archival and records management programs do not have sufficient authority over the procurement and design of systems to insist that new systems satisfy requirements. Several program development projects, designed to enhance the capacity of host institutions to management and preserve electronic records, encountered obstacles because they lacked the authority to act, had insufficient support from top administrators, or did not have access to necessary financial resources.[8] Projects that attempted to extract archival records from existing or inactive information systems confirmed that this approach is time consuming and usually futile.[9] Perhaps the safest conclusion to draw here is that the archival community needs more evidence of the benefits of involvement at the design

[8] These projects are discussed in a summary paper prepared for the Ann Arbor conference. See Hedstrom, M., "Electronic Records Research Issues: A Summary of Recent Research." <http://www.si.umich.edu/e-recs/>

[9] Stout, L., "The Role of University Archives in the Campus Information Environment", *American Archivist* 58:2 (1995): 124-40.

stage, but that there is ample evidence that recovering records from systems that were not designed for recordkeeping is expensive and does not produce satisfactory results.

The Ann Arbor conference offered an unanticipated opportunity to reflect on the processes used to fund and conduct research on electronic records issues during the last five years. For a variety of reasons, research into archival methods is a recent phenomenon in the archival profession. Some archivists have questioned the need for research and suggested that attention to research questions only draws resources away from pressing practical concerns. The structure of the profession has not been conducive to research either. Except for a small teaching faculty, few archivists have the time or the incentives to conduct research. In addition, few archivists are trained in research methods and few archival institutions have the facilities to support research.

The NHPRC research agenda acknowledged the difficulty of conducting basic research within most archival institutions and encouraged multiple strategies for addressing electronic records issues that involved analysis, advocacy, program development, and research. This approach had numerous advantages from the perspective of expediency. Archival institutions with a problem in mind were able to compete for grants that simultaneously would address some aspect of the research agenda and contribute to electronic records program development at their own institutions. Grounded research with constant interaction between the research issues and real world questions helped to keep the research projects closely tied to problems facing the archival community. But this approach to research also has serious limitations, especially in unfamiliar territory where new knowledge is needed. Without consistent methodologies and adequate controls, research results cannot be replicated or easily applied elsewhere. The projects have also suffered from the lack of external review by peers or experts from related disciplines who

were not directly involved in the projects. Finally, the scale of projects needed to make significant breakthroughs in electronic records management and preservation is such that much more substantial projects with much larger budgets are required.

It is interesting to note that the original criteria for projects proposed in the report from the 1991 Research Issues Conference included several factors that would enhance the quality and impact of electronic records research. For example, the working meeting encouraged projects that were suitable for funding by multiple agencies, were multi-disciplinary in conception and execution, would produce usable models and generalizable results, and could be implemented widely. The electronic records research conference, held in Ann Arbor last June, added more criteria to this list. Many conference participants felt that mixing the goals of research and program development in a single project may have compromised achieving either objective. Archival programs, under pressure to get on with their electronic records programs, sometimes found the requirement to contribute to the research agenda a distraction. Ongoing programs, which are subject to a variety of operational, fiscal, and political demands, do not always provide the most appropriate venue for research. The funded projects have contributed partial answers to electronic records research questions and they provided a valuable learning experience. The next step will require a different type of partnership between archival programs and research venues that makes better use of what each as to offer.

What can the archival community do to avoid repeating mistakes that we have made with the first round of electronic records research projects? First, separating research from ongoing program development projects seems like a wise strategy. Encouraging more research in universities and research laboratories also has a number of potential advantages. First, it would help to establish research on electronic records management and digital preservation as legitimate areas of inquiry, and

perhaps open the way for research on other archival issues—use and user requirements, access and retrieval, historical knowledge representation, and visualization—to name just a few opportunities. Second, it would enable archival research to occur in environments that have the facilities, expertise, and other resources needed to support research. Third, moving electronic records research to more appropriate venues would subject both the process and the products to a more rigorous peer review and other formal methods of evaluation. Fourth, it would provide research and educational opportunities for a new generation of archivists who should complete their education better equipped to carry on research. Presumably, this approach would also improve the quality, replicability, and dissemination of research results.

Conducting more electronic records research projects in academic and laboratory settings does not necessarily imply a disjuncture between research and practice. The notion of grounded research is starting to take hold in a number of academic disciplines where real life problems inform research objectives and where a wide variety of organizational, social, cultural, and economic factors are considered in multi-disciplinary research projects. Through vehicles like the electronic records research agenda and repeated requests for answers to electronic records problems, the archival community has provided ample fuel for a vigorous grounded research program. Nor does an academically-based research program relieve ongoing archival programs of responsibility for research. Archival institutions and others struggling with electronic records problems play a critical role in defining the research agenda, testing and refining the results of research, and in evaluation and implementation. A new model of electronic records research would involve more explicit relationships between the investigators, subjects, funders, and consumers of research with clearer roles and responsibilities for each.

In stressing research, I do not want downplay the need for further implementation of the results of recent research. The Pittsburgh Project, for example, produced a high-level model for business acceptable communications which specifies the meta-data needed for reliable records and formalizes production rules for software. Yet only a handful of projects have tested this approach or used the project's functional requirements to analyze the recordkeeping capabilities of systems.[10] Numerous projects developed model policies, guidelines, and standards for record-keeping, but few archival programs are adapting these models to their own administrative and policy environments. What is really needed at this point is a good balance between implementation and evaluation. Archival and records management programs should be encouraged to implement recent research results, carefully monitor both the environmental variables and the implementation process, and report back on the problems they encounter.

A second strategy for advancing electronic records research is to get serious about multi-disciplinary collaboration. The archival community has recognized the value of multi-disciplinary research for some time and several of the NHPRC-funded projects engaged experts from information science, computer science, public policy, and law. Yet research on electronic

[10] Two projects that have used the Pittsburgh requirements during analysis and design are the City of Philadelphia's redesign of its Human Resources Information System <http://www.phila.gov/city/departments/erms/erm.html> and the Indiana University's project to redesign student records and financial accounting systems <http://www.indiana.edu/~libarche/>. The Swedish Pharmaceutical Company, ASTRA, has designed a recordkeeping system for New Drug Applications that addresses reliability, authenticity, and long-term retention. The U.S. Department of Defense has included models and procedures recommended by the UBC project in the redesign of its records management program.

records has not yet benefited fully from the potential for multi-disciplinary collaboration. Archivists need to expand their notions of multi-disciplinary research and extend the range of disciplines with which they collaborate. While there is ample room for more collaboration with computer scientists, policy experts, and lawyers, archivists have forged few ties with such critical fields as communications, management analysis, organizational studies, anthropology, and art and design. Scholars in the humanities, traditionally among archivists' closest allies, are taking a fresh look at the meaning of such basic concepts as memory, evidence, narrative, and interpretation. All of these fields have the potential to inform the theory, methods, and strategies that archivists might use to deal more effectively with electronic records.

Another advantage of a multi-disciplinary perspective is that some of the questions archivists are asking are being addressed by other disciplines. Communications and organizational theorists are examining information flows and authority structures in virtual organizations. Management analysts are studying the organization of work and refining the definition of terms that have crept into the archival vocabulary, such as function, activity, process, and domain. A large body of research on computer-supported cooperative work can help archivists understand what happens to records and information in organizations that use new technologies to enable collaborative endeavors, many of which are time and distance independent. This step of hooking into research that is already underway is also crucial for the archival community to get on top of changes in technology, organization, communications, and documentation.

Although the archival community has made laudable progress toward developing viable solutions to electronic records management, more work is needed before we can determine whether the community has produced a reasonable approach that will carry it from a period of research to one of sustained

implementation and development as David Bearman suggested in his recent summary of electronic records activities.[11] Certainly, implementation and development are needed not only to reap the benefits of recent research, but also to test its limits and challenge some of the unspoken assumption that underlie it. I am not as sanguine as Bearman in suggesting that archivists are about to turn from a research phase to a stage of development and implementation. The recent experience with electronic records research suggests to me that implementation and research need to occur simultaneously, but on parallel tracks with considerable interaction between those archivists, computer scientists, software engineers, business process analysts, and the like who are trying to implement the results of recent research and researchers who are trying to push the boundaries of what we have learned so far.

As I suggested in my 1988 essay, archivists cannot afford the long lead time from posing questions to generating answers. Many of the proposed solutions to electronic records management and preservation have had a short shelf life or they were obsolete by the time they reached a receptive audience. Thinking clearly about the nature of research and moving research into appropriate venues need not detract from an equally important agenda of implementation and development. But as long as the archival community continues to face significant changes in technology, communications, organizational behavior, and scholarship, ongoing research will remain an important ingredient for the long-term solution to preserving electronic records.

[11] D. Bearman, *op. cit.* (1996): 3.

ARTICLE

Increasing the Acceptance of Functional Requirements for Electronic Evidence[1]

Wendy Duff
School of Information Studies, University of Toronto

This article reports on a research study that tested the effect of statements of "literary warrant" on lawyers, auditors and information specialists' evaluations of a set of functional requirements for electronic evidence. It found that legal statements can increase the rating of importance of some of the functional requirements. Its results also provided evidence that differences in subjects professional backgrounds and their computer knowledge can affect the ratings of importance the subjects gave to the functional requirements.

Background

Organizations maintain records to meet the legal, fiscal, and administrative obligations that are dictated by society. Traditionally records managers have emphasized the need to tie records management programs to legal retention requirements. Over the last five years, a number of research projects concerned with the management of electronic records have reaffirmed the importance of grounding one's records programs on a firm legal foundation. For example:

- A survey of New York state agencies' information practices undertaken by the Building Partnership Project found that "program managers were more likely to establish effective recordkeeping systems and practices if .... a clear legal requirement for record retention" existed;[2]
- Bikson and Frinking found that the major obstacle to improving electronic recordkeeping practices in the Dutch government was legislation that did not recognize electronic records as evidence;[3]
- The University of Pittsburgh Project, a three-year study that identified a set of functional requirements for electronic evidence stressed that: "Organizations must comply with the legal and administrative requirements for recordkeeping within the jurisdictions in which they operate, and they must demonstrate awareness of best practices for the industry or business sector to which they belong and the business functions in which they are engaged."[4] These requirements are summarized in Appendix I.

The Pittsburgh Project also postulated that the laws, regulations, case law, information technology standards, auditing standards and best practices promulgated by lawyers, auditors, information specialists, business managers, record managers, and the medical profession could be used to build a strong case for promoting specifications for keeping reliable electronic records. The project staff believed it was important to build this case out

1 This study was completed as part of the requirements for the author's doctoral program. It has benefited immensely from the insightful comments received from members of her doctoral Committee: Richard Cox, Margaret Hedstrom, Edie Rasmussen, and Steven Hirtle. David Bearman first purposed the concept of "literary warrant", and this study has also profited greatly from his wise advice.

2 New York State Archives and Records Administration. Center for Electronic Records, *Building Partnerships for Electronic Recordkeeping: Final Report and Working Papers* (Albany: New York State Archives and Records Administration, Center for Electronic Records, 1995), 10.

3 Bikson, T.K. and E.J. Frinking, *Preserving the Present: Toward Viable Electronic Records* (The Hague: Sdu Publishers, 1993), 15.

4 University of Pittsburgh, School of Information Sciences, "Functional Requirements for Evidence in Recordkeeping." February 1996. <http://www.sis.pitt.edu/~nhprc/>

of a recognition that the specifications for preserving electronic evidence may be undervalued by organizations. They also suggested that other professionals would be more inclined to accept the functional requirements that the Project had identified, if the requirements were tied to laws, regulations and best practices recommended by other professional associations. They stated that "if professionals in our society were made more aware of the functional requirements for recordkeeping as expressed in recommended practices of their own profession (which are themselves grounded in law), they would be more inclined to take responsibility for the adequacy of their recordkeeping practices."[5]

The project team suggested that archivists and/or records managers could use the statements from laws, regulations, standards, etc., as "literary warrant," that is, as proof or justification that organizations and individuals must adhere to the requirements because they are based on practices established by their own professions or industry. The Project suggested that decision-makers would value the requirements more, if the requirements were tied to literary warrant than if they were presented on their own. This article reports on a study that tested that suggestion. Specifically it asked the following questions:

- Does a functional requirement accompanied by literary warrant receive a rating of importance that is significantly different than the rating given a functional requirement by itself?
- Is one type of warrant, that is, warrant drawn from legal, auditing, or information technology literature more influential than others?
- Are there significant differences in the rating of importance of the functional requirements given by different professional groups?

Methodology

The study comprised three different stages. In the first stage of the study, the investigator compiled a list of authoritative sources which relate to professional practices in the law, auditing and information technology fields and dictate requirements for recordkeeping. Nine reviewers in the fields of auditing, law, and information technology (three individuals in each field) rated the authority of the sources from which the warrant was drawn as a check on the credibility of the sources.[6] The authority of each source was scored on a scale of 0-3, with 3 having a great deal of authority and 0 having none.

In the second stage, each authoritative source was scanned for relevant passages that illustrate the functional requirements. Each passage was classified according to the professional group to which the source related and the functional requirement that it supported. Three members of the Pittsburgh project team (Richard Cox, Ken Sochats, and David Bearman) evaluated each passage for its relevance to, and its support for, a functional requirement for evidence.

In the final stage, four research instruments were created. The investigator created the research instruments by first compiling four lists: 1.) just the University of Pittsburgh Project's functional requirements; 2.) the functional requirement augmented with

[5] Bearman, D. et al., "The Warrant for Recordkeeping Requirements", *University of Pittsburgh Recordkeeping Functional Requirements Project: Reports and Working Papers, LIS055/LS94001*, (Pittsburgh: School of Library and Information Science, University of Pittsburgh, 1994), 1.

[6] Three faculty members at the University of Pittsburgh School of Information Sciences evaluated the authority of information technology sources. One internal auditor, one external auditor, and a research director of the Institute of Internal Auditors evaluated the auditing sources. The internal and external auditors had published articles on information systems auditing. Three law professors—two at the University of Pittsburgh and one from another university—evaluated the legal sources.

statements of legal warrant; 3.) the functional requirements augmented with statements of auditing warrant; and finally, 4.) the functional requirements augmented with statements of information technology warrant. The investigator then randomly selected a functional requirement from the first list and assigned it to the first research instrument. A functional requirement and its accompanying auditing warrant was randomly chosen from the second list and assigned to the first research instrument. If the functional requirement already existed in the research instrument (having been taken from another list), a new requirement was selected. A functional requirement and its accompanying legal warrant were randomly chosen from the third list and assigned to the first research instrument. If the functional requirement already existed in the research instrument (having been taken from another list), a new requirement was selected. Then a functional requirement and its accompanying information technology warrant was randomly chosen from the fourth list and assigned to the research instrument. If the functional requirement already existed in the research instrument (having been taken from another list), a new requirement was selected. This process continued until the first research instrument had a complete set of functional requirements (1-20) with five functional requirements being presented on their own, five being accompanied by auditing warrant, five being accompanied by legal warrant, and five being accompanied by information technology warrant.

The second research instrument was created following the same procedure, except that warrants selected for the first research instrument were not included in the selection for second instrument. The creation of third and fourth instruments also followed the same procedure.

Each of the research instruments was presented to one of four different groups of subjects. Each group of subjects consisted of five lawyers, five auditors and five information specialists. Semi-structured interviews were conducted at the subjects' workplace:

- The subjects were asked a few background questions to gather information about their computer and electronic records knowledge.
- Using a script, the investigator briefly described the functional requirements, one at a time, and gave the subjects a piece of paper which contained a functional requirement and any accompanying warrant.
- Where warrant existed, the investigator started by saying that the functional requirement was based on the warrant. Where no warrant existed the subjects were told that the requirement was identified as important by a group of archivists and records mangers.
- The subjects were then asked to rate, on a scale of 1-9 (with 1 being not important at all, 5 being average, and 9 being extremely important) the importance of designing systems that meet this requirement.

The Results

The primary purpose of this study was to test the influence of literary warrant on the scores given to the functional requirements by lawyers, auditors and information specialists. It hypothesized that a person's judgment of a functional requirement would be significantly higher when a functional requirement was accompanied by warrant than when it was presented on its own. As previously noted, the Pittsburgh Project suggested that it was important to build a case for the requirements because on their own the requirements would be undervalued. Before examining the evidence pertaining to warrant, the investigator first sought to determine how important the subjects thought the functional requirements were.

Table 1

Average Scores and Standard Deviations of the Evaluations of the Functional Requirements.

Functional Requirements	Mean	St Dev	Minimum	Maximum
Accurate	8.55	1.00	5.00	9.00
Available	8.45	1.21	3.00	9.00
Consistent	8.05	1.36	3.00	9.00
Inviolate	8.02	1.44	3.00	9.00
Compliant	7.97	1.35	3.00	9.00
Authorized	7.95	1.38	400	9.00
Documented	7.87	1.20	4.00	9.00
Identifiable	7.60	1.40	4.00	9.00
Redactable	7.48	1.75	3.00	9.00
Coherent	7.42	1.59	3.00	9.00
Meaningful	7.35	1.35	4.00	9.00
Renderable	7.28	1.42	4.00	9.00
Implemented	7.23	1.69	2.00	9.00
Auditable	7.17	1.78	3.00	9.00
Comprehensive	7.15	1.68	3.00	9.00
Exportable	7.13	2.01	2.00	9.00
Evidential	6.98	1.68	4.00	9.00
Assigned	6.93	1.81	1.00	9.00
Understandable	6.72	1.68	1.00	9.00
Removable	6.45	1.97	1.00	9.00

The Functional Requirements

The evaluations of the functional requirements with and without warrant were compiled and their means were analyzed. *Table 1* presents the mean scores, their variations and the minimum and maximum scores given to each requirement.

The evaluations of the individual functional requirements varied, with the average score given to *Accurate* being a high of 8.55 with a standard deviation of 1.00, and the average score given to *Removable* being a low of 6.45 with a standard deviation of 1.97.

The frequency and cumulative frequency of each score given to each functional requirement was also tabulated. On the whole, the ratings were high, with only 76 of the 1200 scores being a rating of less than 5. Only four requirements (*Understandable, Removable*, *Exportable*, and *Evidential)* had more than 10% of their evaluations rated as less than 5, on a 9 point scale, while twelve of the twenty requirements received scores of 8 or 9 for more than 50% of their evaluations. *Accurate*, evaluated as the most important requirement, obtained a score of 9, for 78.3% of its evaluations.

The Effect of Warrant

To obtain a comprehensive overview of the average scores given to all the functional requirements without warrant and with legal, auditing and information technology warrant, the means of each functional requirement under each condition were computed. *Table 2* contains the average scores given to each requirement divided into the type of warrant that accompanied it. The condition that received the highest average rating for each functional requirement is marked with a plus sign (+).

Table 2

The Average Ratings Given to the Functional Requirements Accompanied by Different Types of Warrant

Functional Requirement	Without Warrant	With Legal Warrant	With Auditing Warrant	With IT Warrant
Compliant	8.13+	8.07	7.93	7.73
Documented	8.20+	7.4	7.93	7.93
Assigned	6.87	8.07+	5.93	6.87
Implemented	7.0	7.53	7.8+	6.6
Consistent	7.27	8.47+	8.27	8.2
Comprehensive	7.0	7.87+	6.4	7.33
Identifiable	7.53	7.93+	7.87	7.07
Accurate	8.33	8.73+	8.6	8.53
Understandable	6.87+	6.80	6.6	6.6
Meaningful	6.67	7.47	7.47	7.8+
Authorized	7.9	8.87+	7.6	7.5
Inviolate	8.2+	7.47	8.2+	8.2+
Coherent	7.93+	7.33	7.6	6.8
Auditable	6.40	8.00+	7.40	6.87
Removable	5.93	6.73	6.80+	6.33
Exportable	6.87	6.93	8.13+	6.60
Available	8.53	8.73+	8.13	8.4
Renderable	7.07	7.20	7.33	7.53+
Evidential	6.20	7.47+	7.00	7.27
Redactable	7.27	7.20	8.20+	7.27

Of the twenty functional requirements, the highest average score for nine of the requirements was attained when they were accompanied by legal warrant. Four functional requirements received their highest average scores when they were accompanied by auditing warrant and another four requirements received the highest average scores when they were accompanied by no warrant. Only two functional requirements received their highest average scores when they were accompanied by information technology warrant. There was one tie among the highest average scores that one functional requirement received. *Inviolate* received the same score when it was accompanied by auditing warrant, when it was accompanied by information technology warrant, and when it was presented without warrant.

Seven functional requirements received their lowest average scores when they were accompanied by no warrant. The lowest scores for six functional requirements were attained when they were accompanied by information technology warrant. Three requirements received their lowest average score when they were accompanied by auditing warrant and another three received their lowest average scores when they were accompanied by legal warrant. There was one tie between the lowest average score that a functional requirement attained when it was accompanied by information technology and auditing warrant.

To discover if the scores given to the functional requirements accompanied by different types of warrant were significantly different, an Analysis of Variance including the Scheffe test was conducted. Only two requirements, *Assigned* and *Authorized* had any significant differences in their evaluations, and both received their highest mean rating when they were accompanied by legal warrant.

An Analysis of Variance with repeated measures was conducted to obtain a better understanding of the overall effect of warrant. The four groups of subjects were analyzed separately

and the data provided by each group of subjects indicated a strong relationship or synergy between the warrant and the functional requirement, which was significant at the .01 level. In some cases the presence of warrant can make a significant difference, but in other cases it is the strong relationship between a particular piece of warrant and a specific functional requirement that is significant.

Professional Differences

To discover if the different professional groups were affected by the warrant, the data were divided by professional group and an Analysis of Variance was conducted. This method of analysis resulted in extremely small samples (only five subjects in each category). With these small sample sizes, only lawyers showed any significant differences in the ratings they assigned the requirements accompanied by different types of warrant. The presence of warrant significantly affected the lawyers' evaluations of four of the twenty requirements, *Assigned, Implemented, Authorized* and *Auditable*. The presence of warrant did not cause any statistically significant differences in the evaluations provided by the auditors or information specialists.

Persuasion research suggests that when the cognitive effort is too great to interpret a message, people will use simple decision rules that depend upon persuasion cues, e.g., source credibility, with message validity.[7] Research also suggests that people with a strong working-knowledge are less likely to be persuaded by heuristics or simple persuasion cues, such as the credibility of a message, than people with less working-knowledge.[8] The lawyers in the study had far less knowledge of computers and electronic records, and therefore they probably would have needed greater cognitive effort to interpret the functional requirements than the information specialists or the auditors. The finding that warrant was more effective in influencing the opinions of lawyers than information specialists or auditors is not surprising considering the conclusions of previous persuasion research.

An Analysis of Variance was performed to discover if there were significant differences in the scores given by the different professional groups. The differences in the rating given to the requirements by the different professional groups were statistically significant, at the .05 level, for only one functional requirement, *Comprehensive*. The highest average ratings were provided by auditors. *Authorized* showed a strong, but not quite significant difference, in the scores given by the different professional groups, with information specialists giving the highest average score.[9]

To obtain a overview of the average scores provided by the different professional groups, the means of each functional requirement given by each professional group were computed. *Table 3* contains the average scores given to each requirement divided into the professional group that provided it.

[7] Petty, R.E. and J.T. Cacioppo, *Communication and Persuasion: Central and Peripheral Routes to Attitude Change* (New York: Springler-Verlag, 1986); and Shelly Chaiken, "Heuristic Versus Systematic Information Processing and the Use of Source Versus Message Cues in Persuasion", *Journal of Personality and Social Psychology* 39:5 (1980): 752-766.

[8] Wood, W. and B. Stagner, "Why Are Some People Easier to Influence Than Others", in S. Shavitt and T.C. Brock (eds.), *Persuasion: Psychological Insights and Perspectives* (Boston: Allyn and Bacon, 1994), 149-174.

[9] *Authorized* has been included as almost significant even though the probability of .0553 is slightly higher than the .05 level.

Table 3

The Average Ratings Given to the Functional Requirements by the Professional Groups

Functional Requirement	Information Specialists	Lawyers	Auditors
Compliant	8.00	7.70	8.20
Documented	8.25+	7.65	7.70
Assigned	6.95	6.60	7.25
Implemented	7.60+	7.05	7.05
Consistent	8.30	7.50	8.35
Comprehensive	7.05	6.30	8.10
Identifiable	8.00+	7.30	7.50
Accurate	8.55	8.60+	8.50
Understandable	7.05+	7.00	6.10
Meaningful	7.65+	7.25	7.15
Authorized	8.25+	7.35	8.20
Inviolate	8.15	7.65	8.25
Coherent	7.40	7.40	7.45
Auditable	7.25	6.60	7.65
Removable	6.45	6.50+	6.40
Exportable	7.45+	7.35	6.60
Available	8.15	8.60+	8.60
Renderable	7.50+	7.15	7.20
Evidential	7.30+	6.50	7.15
Redactable	7.75+	7.60	7.10

The professional group that received the highest average rating for each functional requirement is marked with a plus sign (+).

Of the twenty functional requirements, ten received their highest mean scores from information specialists, while seven received their highest mean scores from auditors. Lawyers provided the highest mean scores for only four functional requirements. There was one tie between the highest average scores that a functional requirement received when it was evaluated by lawyers and when it was evaluated by information specialists.

On the other hand, lawyers provided the lowest mean scores for eleven of the twenty evaluations. Only one functional requirement received its lowest mean score from information specialists. There was one tie between the lowest average scores that one functional requirement received when it was evaluated by lawyers and when it was evaluated by auditors. There was also one tie between the lowest average scores that another functional requirement received when it was evaluated by lawyers and when it was evaluated by information specialists.

On the whole, information specialists and auditors seemed to have more appreciation for the functional requirements than the lawyers in the study. Lawyers provided the highest average score for only two requirements and gave the lowest average score to eleven functional requirements. Perhaps individuals with a high degree of computer knowledge, like the information specialists and auditors in this study, can appreciate more fully the importance of the requirements.

The ranking of the functional requirements also suggest that the different professional groups value the requirements differently. Auditors rated *Comprehensive* as the seventh most important requirement, while information specialist ranked it as the eighteenth most important requirement, and lawyers ranked it last. Information specialists ranked *Authorized* as the second most

important requirement, auditors rated it as fifth, and lawyers ranked it as the ninth most important requirement. *Redactable*, however was ranked as the sixth most important requirement by lawyers, the ninth most important requirement by information specialists, and the fifteenth most important requirement by auditors.

Auditors may rate *Comprehensive* as more important than the other professional groups because their training and experience is related often to auditing financial records.[10] The concept that a system should capture all financial transactions is well established. The auditors in this study may have equated the requirement, *Comprehensive,* to the capturing of financial transactions, rather than all transactions. The professional background of a subject also strongly affected the scores given to *Authorized*, with information specialists ranking this requirement higher than lawyers or auditors. Many of the techniques for authorizing data, such as digital signatures or passwords, have been developed to fulfill requirements for data and systems security. Perhaps information specialists understand the close connection between this requirement and security measures, and therefore ranked it as more important than the other professional groups. Lawyers ranked *Redactable* higher than the other professional groups

[10] Many of the standards and principles that regulate an auditor's professional activities are promulgated by Accounting Associations. For example, the Financial Accounting Standards Board, the American Institute of Certified Public Accountants (AICPA), the Auditing Standards Board and the AICPA Quality Control Standards Committee all issue standards that guide the work of auditors. Many of these standards highlight the requirements of auditing financial statements. For example, the American Institute of Certified Public Accountants, Codification of Statements on Auditing Standards, Numbers 1 to 73, AU Section 326.05 : Evidential Matter, Nature of Assertions, 1994, points out that "assertions about completeness deal with whether all transactions and accounts that should be presented in the financial statements are so included."

suggesting that lawyers have a greater appreciation for the need to mask confidential information.

Effect of the Subjects' Background

The differences in the ratings given by the three professional groups may be due to differences in professional perspective, or they may have arisen from differences in the subjects' background. To discover if there was a relationship between these variables and the evaluations given to the functional requirements, data on the subjects' background were collected and correlation analysis conducted. Correlations tended to be small and not significant, with the exception of the subjects' knowledge of, and experience with, computers. The ratings given to the requirements *Comprehensive*, *Authorized* and *Auditable* correlated positively with the subjects' computer knowledge. The ratings given to *Documented* correlated positively with the years the subjects had used computers.

Subjects may gain a greater appreciation for the importance of documenting systems, capturing all records, maintaining audit trails of a record's use, and ensuring only authorized individuals create records, as they gain knowledge of, and experience with, computers. This does not suggest that people will suddenly understand the need for reliable evidence as they become computer literate. Rather, it suggests that some of the requirements needed to develop trustworthy information systems, such as having systems documentation and ensuring adequate security measures, are similar to some of the requirements for having reliable evidence.

Summary

Warrant can successfully increase the acceptance of the functional requirements in some cases. The scores given to certain functional requirements, in particular *Assigned* and

Authorized, were affected significantly by the presence of legal warrant. Legal warrant appears to have the greatest influence, and lawyers appear to be more influenced by warrant than information specialists or auditors. What is not known is whether legal warrant had its greatest influence because legal sources have the greatest authority, or whether legal warrant had the greatest influence because lawyers, on the whole, were the most influenced by it.

Analyzing the functional requirements as a whole, rather than individually, provides evidence that there is a significant relationship between the functional requirements and the warrant. Developing a warrant for the functional requirements is important, although it may be more important for increasing the acceptance of some requirements than others, and it may be more effective in influencing the opinion of some professional groups than others. For example, it may have a greater effect on influencing the opinions of people with less computer knowledge. Although warrant may not significantly increase the acceptance of individual functional requirements, its presence may have enhanced the credibility of all the requirements. This study supports, in part, the belief that warrant may be an important tool that archivists and records managers can use to influence the design of recordkeeping systems and ensure that electronic records are captured and maintained in the future. More research is needed to verify this conclusion.

Warrant is not the only solution needed to ensure the requirements for reliable evidence become accepted and incorporated into the design of systems. The subject's computer knowledge and experience showed a mild positive correlation with the rating of importance of some functional requirements and the professional background strongly affected the rating of two requirements. The study also showed that a strong working-knowledge of computers may decrease the influence of warrant. There is a strong relationship, or synergy, between the functional requirements and the warrant which needs further study. Further research on warrant, including the effect of different types of warrant on other professional groups and the effect of the relationship between the warrant and the requirements, needs to be conducted. When this research is undertaken the results may point to new strategies needed to ensure that archivists and records managers become involved in the design of record-keeping systems that capture and preserve electronic records.

APPENDIX I

Functional Requirements for Recordkeeping

As noted in the introduction, the University of Pittsburgh responded to the NHPRC's electronic records research agenda by conducting a research project to develop a set of functional requirements for recordkeeping. These requirements are system independent and could be implemented in either a manual, electronic or hybrid system.

The nineteen requirements for recordkeeping are grouped into three different categories:

- requirements that relate to the organisation—labeled Compliant organisation;
- requirements reflecting specifications for recordkeeping systems—classified as Accountable recordkeeping systems;
- requirements that relate to Records—grouped under three sub-categories: Captured Records; Maintained Records, and Usable Records.

Compliant Organization

1. Compliant

Accountable Recordkeeping System

2. Responsible
3. Implemented
4 Consistent

Captured Records

5. Comprehensive
6. Identifiable
7. Complete
   7a. Accurate
   7b. Understandable
   7c. Meaningful
8. Authorized

Maintained Records

9. Preserved
   9a. Inviolate
   9b. Coherent
   9c. Auditable
10. Removable

Usable Records

11. Exportable
12. Accessible
    12a. Available
    12b. Renderable
    12c. Evidential
13. Redactable

Organization: Compliant[11]

1. Compliant

Organizations must comply with the legal and administrative requirements for recordkeeping within the jurisdictions in which they operate, and they must demonstrate awareness of best practices for the industry or business sector to which they belong and the business functions in which they are engaged.

1a. External recordkeeping requirements are known.

1a1. Laws of jurisdiction with authority over the record creating organizations are known.

1a2. Regulatory issuances of entities with administrative authority over the record creating organizations are known.

1a3. Best practices of recordkeeping established by professional and business organizations within the industry and business functions of the organization are known.

1b. Records created by organizational business transactions which are governed by external recordkeeping requirements are linked to an internal retention rule referencing the documented law, regulation, or statement of best practice.

1c. Laws, regulations, and statements of best practice with requirements for recordkeeping are tracked so that changes

[11] To ensure an organisation's system meets its regulatory obligations, an organisation must first identify all relevant legal and administrative requirements with which it must comply. The organisation must identify the records that it must keep, the time period for which they are needed, and how to preserve them. The system must create linkages between the records created to meet these requirements and rules that control their retention and/or disposition. The organisation must monitor the laws, regulations and best practices in their regulatory environment and incorporate any changes to the regulations into their systems.

to them are reflected in updated internal recordkeeping instructions.

Recordkeeping Systems: Accountable[12]

2. Responsible

Recordkeeping systems must have accurately documented policies, assigned responsibilities, and formal methodologies for their management.

2a. System policies and procedures are written and changes to them are maintained and current.

2b. A person or office is designated in writing as responsible for satisfying recordkeeping requirements in each system.

2c. System management methods are defined for all routine tasks.

2d. System management methods are defined for events in which the primary system fails.

3. Implemented

Recordkeeping systems must be employed at all times in the normal course of business.

3a. Business transactions are conducted only through the documented recordkeeping system and its documented exception procedures.

3b. No records can be created in the recordkeeping systems except through execution of a business transaction.

[12] The environment in which records reside can either increase or decrease their reliability and trustworthiness. The courts bestow a high degree of trust in records that are "kept in the regular course of business activity ...as shown by the testimony of the custodian or other qualified witness, unless the source of information or the method or circumstances of preparation indicate lack of trustworthiness." The admissibility of records depends upon testimony that verifies the integrity and reliability of the recordkeeping system that controlled them. Therefore, the second group of requirements delineate specifications for the recordkeeping system and are labelled *Responsible, Implemented,* and *Consistent.*

3c. Recordkeeping systems and/or documented exception procedures can be demonstrated to have been operating at all times.

4. Consistent

Recordkeeping systems must process information in a fashion that assures that the records they create are credible.

4a. Identical data processes permitted by the system must produce identical outcomes regardless of the conditions under which they are executed.

4b. Results of executing systems logic are demonstrable outside the system.

4c. All operational failures to execute instructions are reported by the system.

4d. In the event of system failures, processes under way are recovered and re-executed.

Records: Captured[13]

5. Comprehensive

Records must be created for all business transactions.

5a. Communications in the conduct of business between two people, between a person and a store of information available to others, and between a source of information and a person, all generate a record.

5b. Data interchanged within and between computers under the control of software employed in the conduct of business creates a record when the consequence of the data processing function is to modify records subsequently employed by people in the conduct of business.

6. Identifiable

Records must be bounded by linkage to a transaction which used all the data in the record and only that data.

[13] The third category of requirements specify characteristics that records must have, and they are arranged into three subgroups: *Captured Records, Maintained Records,* and *Usable Records.*

6a. There exists a discrete record, representing the sum of all data associated with a business transaction.

6b. All data in the record belongs to the same transaction.

6c. Each record is uniquely identified.

7. Complete

Records must contain the content, structure, and context generated by the transaction they document.

7a. Accurate: The content of records must be quality controlled at input to ensure that information in the system correctly reflects what was communicated in the transaction.

7a1. Data capture practices and system functions ensure that source data is exactly replicated by system or corrected to reflect values established in system authority files.

7b. Understandable: The relationship between elements of information content must be represented in a way that supports their intended meaning.

7b1. Meaning conveyed by presentation of data are retained or represented.

7b2. System defined views or permissions are retained and the effects are reflected in the record represented.

7b3. Logical relations defined across physical records are retained or represented.

7b4. Software functionality invoked by data values in the content of the record are supported or represented.

7c. Meaningful: The contextual linkages of records must carry information necessary to understand correctly the transactions that created and used them.

7c1. The business rules for transactions, which minimally locate the transaction within a business function, are captured.

7c2. A representation of the source and time of the transaction which generated a record is captured.

7c3. Links between transactions which comprised a single logical business activity are captured.

8. Authorized

An authorized records creator must have originated all records.

8a. All records have creators which are documented.

8b. Records creators must have been authorized to engage in the business that generated the record.

Records: Maintained[14]

9. Preserved

Records must continue to reflect content, structure, and context within any systems by which the records are retained over time.

9a. Inviolate: Records are protected from accidental or intended damage or destruction and from any modification.

9a1. No data within a record may be deleted, altered, or lost once the transaction which generated it has occurred.

9b. Coherent: The information content and structure of records must be retained in reconstructible relations.

9b1. If records are migrated to new software environments, content, structure, and context information must be linked to software functionality that preserves their executable connections or representations of their relations must enable humans to

[14] Once records are captured, they must be maintained over time. Therefore, the functional requirements for recordkeeping contain specifications for migrating records to new hardware and software environments. To be preserved, records must maintain their content, structure and context regardless of the software and hardware controls under which they exist. Therefore, the functional requirement *Preserved* is divided into three separate sub-requirements: *Inviolate, Coherent,* and *Auditable.*

reconstruct the relations that pertained in the original software environment.

9b2. Logical record boundaries must be preserved regardless of physical representations.

9c. Auditable: Record context represents all processes in which records participated.

9c1. All uses of records are transactions.

9c2. Transactions which index, classify, schedule, file, view, copy, distribute, or move a record without altering it are documented by audit trails attached to the original record.

9c3. Transactions which execute a records disposition instruction, whether for retention or destruction, are documented by audit trails to the original record.

10. Removable

Records content and structure supporting the meaning of content must be deletable.

10a. Authority for deletion of record content and structure exists.

10b. Deletion transactions are documented as audit trails.

10c. Deletion transactions remove the content and structural information of records without removing audit trails reflecting context.

Records: Usable

11. Exportable

It must be possible to transmit records to other systems without loss of information.

11a. Exporting protocols should be reversible.

11b. Functionality should be represented in a fashion that produces the same result in the target system as in the originating environment.

12. Accessible

It must be possible to output record content, structure, and context.

12a. Available: Records must be available.

12b. Renderable: Records must display, print, or be abstractly represented as they originally appeared at the time of creation and initial receipt.

12b1. The structure of data in a record must appear to subsequent users as it appeared to the recipient of the record in the original transaction or a human meaningful representation of that original rendering should accompany the presentation of the original context.

12c. Evidential: Record's representations must reflect the context of the creation and use of the records.

13. Redactable: Records must be masked when it is necessary to deliver censored copies and the version as released must be documented in a linked transaction.

13a. The release of redacted versions of a record is a discrete business transaction.

13b. The fact of the release of a redacted version of a record is an auditable use of the original record and therefore results in creation of an audit trail with a link to the transaction which released the redaction.

ARTICLE

Harnessing The Potential: Effective Access To Cultural Heritage Information

Lyn Elliott Sherwood
Canadian Heritage Information Network

Introduction

This article had its origins in the author's personal reflections on the practical implications of the new mission adopted by the Canadian Heritage Information Network (CHIN): "to broker effective access to Canadian and international heritage information for public education and enjoyment and for the collective benefit of the Canadian museum community". In attempting to fulfil the mission, we play multiple roles including, but not limited to: content provider, value-added gateway or premium service provider, consultant on new approaches to information management, participant in standards development and promoter of standards. These roles must evolve in the context of an environment in which all the targets move rapidly and they must exist in relation to the roles played by other stakeholders including our member institutions, generic network utilities and services, other organizations which are active in developing solutions for networked cultural heritage and the audiences for heritage information. Attempting to define appropriate roles for CHIN, therefore, means understanding how these other roles may also evolve.[1]

A key issue for us and for our members is how to ensure that audiences are well-served in their quest for cultural heritage content and, in meeting audience needs, where to apportion the necessary investment along the information and access "food chain". This article surveys current issues in the provision of effective access to networked information in order to assess the implications of these issues for the evolving roles of the various stakeholders identified above. In so doing, it highlights some existing approaches and promising advances, suggests areas in which collaborative efforts involving not only members of the cultural heritage community might prove fruitful, and identifies a number of outstanding questions where further research and analysis will be necessary.

Most cultural heritage information available in today's networked environment serves a limited role in promoting the institutions themselves. As a community, we are only at the earliest stages of providing real content and of grappling with the issues buried in the concept of "effective access". In interpreting our mission, we are assuming at CHIN that providing effective access will rest not only on the availability of a critical mass of quality content in digital form but also on recognition that the definition of effectiveness will vary in a diverse audience, on strategies to manage information as a durable and flexible asset and on appropriate functionality to facilitate resource discovery and retrieval in the context of what has been rightly termed an "information explosion". Fulfilling these conditions will require ongoing investment and an enhanced understanding of the economic models which will shape the investment.

[1] I am indebted to discussions with many individuals during the preparation of this article and would like particularly to thank John Perkins and Paul Evan Peters as well as other members of the CIMI Executive Committee, managers and staff at CHIN, Howard Besser, David Bearman, and Guy Herman.

Context

In the context of rapid network evolution and in seeking to navigate through the plethora of material to which we now have potential access, many of us have experienced the learning curve which Nicky Ferguson of the University of Bristol has characterized as scepticism, followed by excitement, followed in short order by frustration.[2] It has become clear that, without the mechanisms to manage extraordinary volumes of information, we will not be able to harness the true potential of global networks for scholarship, business purposes, lifelong learning and entertainment. Equally clear is that the solutions available to us today will not be sufficiently scalable to constitute an adequate response.

Solutions for the networked environment will need to take account of the fact that users of differing ages and academic stages may have widely divergent needs and expectations. A common metaphor for the networked environment is a vast digital library. Within that definition, however, rest many assumptions about the experience of the library user. For some, the benefit is a vast array of information on all subjects, from which material relevant to a current interest may be selected. These users may locate items simply by wandering the shelves until something catches their attention. For others, the benefit of the library is the organization of holdings which allows them to focus on the range of information available on a given subject. Some may wish only to seek a specific title, others to consult reference works such as encyclopedias which provide useful summaries of complex topics. Some researchers find the most useful attribute of the specific title to be the bibliography or the index if the objective is to either broaden or narrow the search.

[2] Ferguson, N., presentation at the March 1996 meeting of the Coalition for Networked Information, Washington, D.C.

Irrespective of the needs, expectations or skills a user may bring to the experience, in today's networked information environment we are very far from achieving what the Coalition for Networked Information's White Paper on Networked Information Discovery and Retrieval (NIDR) defines as a digital library: "an organized, selected or managed body of information".[3] While many of the search engines operate by classifying sites, some estimates indicate that fewer than 50% of existing sites have in fact been indexed or catalogued. An increasing number of network resources or services (particularly non-textual materials) may actually be invisible to existing Web indexing utilities. The cataloguing itself may be based on arbitrary knowledge classification systems which fail to recognize the multiplicity of taxonomic relationships among domains of knowledge. There is no selection or overall management: there are no restrictions as to who may create a Web site and research on a given topic will yield anything from personal homepages to academic, commercial, not-for-profit and government information sources with no inherent means of determining reliability or currency. Sites are closed or change their URLs and existing pointers cease to function.

In the present environment, a very high percentage of the cost of information discovery and retrieval is borne by the end-user in the form of a heavy investment of time with an increasingly low rate of return as measured by satisfaction with the results. At the same time, generic service providers are making significant investments in the operation and enhancement of search engines while specialized or premium service providers are investing in the development of subject-specific agents designed to exploit the diverse strengths (or overcome the diverse weaknesses) of the

[3] Lynch, C. et al., draft white paper on Networked Information Discovery and Retrieval, Coalition for Networked Information. <http://www.cni.org>

more generic search engines.[4] Other service providers are developing alternate approaches to enhancing the location and retrieval of information in specific knowledge domains.

On the other side of the coin, holders of information such as museums, archives, libraries and universities are beginning to make significant investments in developing resources in digital form. In a series of brainstorming sessions which CHIN held in 1995 with Canadian museum directors, participants emphasized the potential of the Internet and other digital formats to assist them in their missions to make their collections better known, to reach new audiences and to generate new revenues to offset reductions in public funding.[5] Their preoccupations included not only access to the information for which they are responsible, but also maximizing the return on their investments by enabling the use of information for a variety of purposes and its survival through successive generations of supporting technologies. Despite this concern with the long-term protection of investment, however, most current digitization projects do not incorporate mechanisms which will facilitate the re-use of material within the

[4] For interesting overviews of development work for search engines, see Steinberg, S.G., *Wired* (1996), 108-114 and 172-181; see also Marchionini, G., "Resource Search and Discovery", *Research Agenda for Networked Cultural Heritage* (The Getty Art History Information Program [Getty Information Institute], 1996), 35-40.

[5] For more detailed discussions of these brainstorming sessions and their implications, see "Canadian Heritage Information Network", a paper presented by the author at the Science Museum, London, 10 May 1995; see also Bearman, D., "Information Strategies and Structures for Electronic Museums", in A. Fahy and Dr. W. Sudbury (eds), *Information: The Hidden Resource, Museums and the Internet, Proceedings of the Seventh International Conference of the MDA, 1995,* (MDA, 1995), 5-22. Copies of the summary report of the sessions may be obtained from CHIN. <service@chin.gc.ca>

organization, dynamic linking of this content with material from diverse sources or migration to new platforms.

Generic Resource Discovery

The sheer volume of information available on virtually any topic is beginning to make the task of information discovery increasingly vexatious. The browser who wishes to "wander the shelves" in search of random success may be satisfied but the person who seeks a more specific result will be frustrated. Search results of 25,000 items are useful to only the most dedicated researcher. Equally, nil results carry no certainty that nothing exists; they merely reflect the fact that a given search engine has failed to locate relevant material.

When CHIN held its brainstorming sessions with museums, there was a certain level of optimism about the reliability of the search engines available through any standard Internet browser and debate as to whether collective approaches were necessary to assist audiences to locate Canadian museum sites. It remains true that if a searcher wishes to reach a specific site, he or she can use the existing utilities relatively effectively. Reliance on this type of access places the onus on the institution to market its presence and clearly favours the larger and better-known institutions. Eighteen months after this discussion (an eon in Internet life) no single search engine, however, would yield a comprehensive list of all Canadian museums with Web sites. Few casual browsers will take the trouble to do successive searches using a variety of engines. The visibility of an individual institution is thus dependent on a series of random factors which may lie outside its control.

Much of the current development work in search engines appears to be focussed on the issue of the classification of material to support resource discovery. Enhanced indexing and cataloguing approaches are being developed based on both

human effort to refine and enhance taxonomic schemes and computer-assisted analysis to resolve what Steinberg terms the issues of "synonymy and homonymy".[6] Utilities such as thesauri are being incorporated into the search architectures as additional solutions to the problem of synonymy, more sophisticated algorithms to support features such as relevance ranking are being implemented, and various artificial intelligence applications are being tested to resolve the problem of homonymy through contextual analysis. Many of these developments may in fact increase the required investment on the part of the user by yielding expanded search results. The value of the information provider's investment in creating a resource may actually be reduced as a result of the increased competition for audience attention.

Value-Added Access

Generic search engines are unlikely to provide the combination of comprehensiveness, intellectual sophistication in a given subject area (e.g., incorporation of domain-specific thesauri or more sophisticated AI applications) and assessment of the authoritativeness of sources needed to satisfy scholarly requirements for resource discovery. Nor will they necessarily support services which may be insisted upon by commercial audiences (e.g., the storage of user search profiles and "new information only" services , specialized user interfaces to serve particular needs or transactional capabilities). Premium service providers who supply value-added access are now emerging, inevitably further advanced in commercial or profitable fields such as medicine, law, science and engineering than in the humanities.

6 Steinberg, *op. cit.*, (1996), 176.

In the humanities, various interim strategies, usually community-based, are being adopted to assist users to navigate through the growing maze of available resources. One of the core principles on which CHIN founded its new strategic directions in 1995 was that approaches which provided member museums with a collective presence in global networks provided a greater opportunity for them to realize their individual objectives of making their collections better known, reaching new audiences and generating new revenues than could be achieved through independent institutional efforts.

The tactics originally pursued by CHIN to increase the effectiveness of audience access paralleled those of many other public sector and academic organizations: hot links to sites operated by our members and to other relevant sources of information. Although additional approaches have now been adopted, CHIN still maintains such lists, compiled through incidental research on the part of staff, professional contacts, courtesy exchanges of URLS and other similar means. Because the lists and links are maintained manually, they inevitably suffer from continual problems of currency as sites are closed or change addresses.

An interesting advance on the relatively random creation of lists and links is represented by projects such as the Social Science Information Gateway (SOSIG), operated by the Centre for Computing in the Social Sciences at the University of Bristol.[7] Relevant sites are selected after an evaluation process involving staff research and input from the academic community and other users. Sites are classified (based on Universal Decimal Classification schema) and catalogued with keywords, resource type identifiers, contact information and a summary description. As an initiative with extensive community involvement, this

7 See SOSIG Web site for further information. <http://www.sosig.ac.uk>

strategy holds promise in addressing concerns about the authoritativeness and relevance of sources.

Discussion of the SOSIG project at the March 1996 meeting of the Coalition for Networked Information in Washington, D.C., did highlight the issue of scalability. Participants argued that, as the number of catalogued sites increases, researchers will seek more refined results from their searches, not simply the site but specific information within the site or in multiple sites. To address this issue, the Arts and Humanities Data Service in the U.K.[8] envisions a three-tiered navigational approach: general resource discovery at the umbrella site; more detailed descriptions and finding aids at the level of the domain-specific service providers; and "implementation level" information at the level of individual data sets.

Each of the above strategies for providing value-added access is relatively labour-intensive, with investment either on the part of a service centre or the content creator (in the case of the third level in the AHDS tiers). Increasing attention is being paid to the prospect of using agents to develop comprehensive indexes relevant to particular subjects.[9] In addition to intelligent agents developed by specific premium service providers, there are now several commercial products which can undertake multiple-engine searches and which embed varying levels of intelligence. The results of these agent-based searches, however, will still require manual verification to ensure both a reliably-parsed index and the relevance of identified sites. Nevertheless, these agents offer the opportunity to diminish investment in repetitive manual resource discovery processes. For value-added service providers in non-commercial domains such as cultural heritage, the availability of such agents may offer the opportunity to replace investment in software development with increased investment of intellectual capital in developing refined searches and capturing the results for the benefit of the particular community which they serve.

A question which will ultimately need to be answered, however, is whether professional communities or other users of the resources will be willing to finance these service activities, most of which have been initiated with temporary funding. SOSIG, for example, has been funded by the Electronic Libraries Program. The recently-launched American Arts and Letters Network (AALN), which intends to facilitate "access to digital resources in the arts and humanities considered integral to most levels of teaching, scholarship, and learning", has been funded by the Andrew W. Mellon Foundation.[10] The Arts and Humanities Data Service received funding for three years from the Joint Information Systems Committee of the Higher Education Funding Councils.

Index or Content?

To date, even a refined search-engine result will leave the user at the level of general index, with the further requirement to review sites sequentially and, within many sites, to spend time searching for specifically relevant material. A longer-term

[8] Information on the navigational approach proposed by AHDS was provided by Daniel Greenstein, Executive Director for the project, at the Digital Resources for the Humanities conference (Oxford, 1-3 July 1996). For information on AHDS, see their web site. <http://www.kcl.ac.uk/projects/ahds/index.html>

[9] See the Harvest project web site for a description of an adaptable agent. <http://harvest.cs.colorado.edu/>

[10] The project is a collaborative effort by the American Council of Learned Societies, the Coalition for Networked Information and Vassar College and is funded by the Andrew W. Mellon Foundation. For further information, see the project web site. <http://www.aaln.org>

question remains to be answered: will users be satisfied with results which consist of lists or pointers to sources which must be reviewed sequentially or will there be an expectation of substantive and precise results drawn from a range of sources (a merging in some respects of the discovery and retrieval processes)? Drawing an example from CHIN's environment, a tourist might prefer to pose a single question, "what special museum exhibits can I visit in Vancouver next week?" than to search sequentially through a series of Web sites. A long-term objective for CHIN is to achieve effective retrieval of specific information across multiple sources with an integrated presentation of results.

As an interim strategy pending the development of robust solutions to this end, we have developed, in the *Guide to Canadian Museums and Galleries* and the *Heritage Forum*,[11] hybrid applications which provide for integrated searching on specific information elements and sequential searching on related content. Information in searchable categories is held centrally (fed by Internet form or FTP to a document-management database engine) for consultation by audiences. Hypertext links from the applications to related sources of information in a widely distributed environment allow audiences to obtain more comprehensive information.

While the two applications identified in the previous paragraph, offer an immediate benefit in terms of integrated results, they are currently limited by the need to define the specific questions which may be posed and by the requirement for participating content providers to choose to submit appropriate material. In the latter context, participants may actually need to increase their level of investment in the preparation of

[11] These applications are located on the CHIN Web site. <http://www.chin.gc.ca>

information if similar material is required in different form for local applications. In seeking to move from the hybrid approach embodied in these applications to a truly distributed information environment, at issue is whether audience requirements for more substantive or integrated responses to queries can, under any scenario, be satisfied without at least some investment in the solution on the part of content providers. Whether such investment will be made depends, of course, on the perceived return.

Content Providers

As part of its effort to realize the long-term objective of searching and retrieving across multiple sources, CHIN is a sponsoring member of the Consortium for the Computer Interchange of Museum Information (CIMI). The Consortium, an international initiative which now includes 16 organizations,[12] is developing and promoting standards for structuring and retrieving cultural heritage information in digital form with a view to enabling integrated multiple media resources. Particular attention is being paid to the development of a museum-specific document-type definition (DTD) for markup of text in Standard Generalized Markup Language (SGML) and an attribute set for museum information for the Z39.50 exchange protocol.

[12] CIMI members in November 1996 included: the Getty Information Institute; CHIN; the Research Libraries Group; the Museum Computer Network; the National Museum of American Art; the National Gallery of Art, Washington; the University of California Berkley, Museum Informatics Project; the University of California Division of Library Automation; the Coalition for Networked Information; the Canadian Museum of Civilization; Corbis Corporation; the Museum Documentation Association (U.K.); the Victoria and Albert Museum; the RAMA Consortium; the Chicago Historical Society; and de Montfort University. See CIMI Web site for details of CIMI's work plan and reports on projects. <http://www.nstn.ca/cimi>

An extremely important issue which is surfacing in the context of CIMI's work on SGML is the challenge of developing a business case for its implementation by institutions. The lack of tools to facilitate both the markup process and the retrieval of encoded material is an immediate practical impediment given the labour-intensive nature of the task. A more serious consideration is whether museums will consider their information or knowledge as a strategic asset which must be appropriately managed for long-term benefit. The answer will lie in museums' definition of the role they wish to assume in a knowledge economy: key sources of content, either through the provision of "raw" information elements or as the authors of meaningful knowledge products delivered via networks to remote audiences; or bit-players in the networked environment, continuing to use the resource primarily for marketing or promotional purposes but placing primary emphasis on the more traditional roles of managing physical collections and direct visitor experiences.

If museums do wish to become key players in the knowledge economy, they will need to follow the lead of organizations in other sectors in tackling the issues of comprehensive information management. The structured databases on which most museum investment has previously been centred have generally been developed only for internal use and may not be easily adaptable for public access (in terms of both content and design).[13] In the current investment in digital information destined for the public via the Web, material is generally being developed for a single application and its re-use in subsequent applications will require extensive manipulation and preparation. The investment in research, writing and digitization cannot, therefore, be easily leveraged through successive uses. Nor is attention necessarily paid to linking various types of information (catalogue, narrative, images, audio and video resources) which may be held in various locations and formats across the organization or, indeed, across several organizations.

If information is to be re-used or linked with resources created at different times by different individuals, efficient mechanisms will be needed both to offset a high degree of variability in the ways in which knowledge is represented by the diverse "authors" and to provide for efficient discovery and retrieval of the information. The search-engine/service provider enhancements alluded to above (better indexing, thesauri etc.) will go some way towards addressing these requirements for textual information. Inclusion of authorities such as the Art and Architecture Thesaurus or the Union List of Artists' Names in a search architecture, for example, might enhance the retrieval of cultural heritage information (as demonstrated in the Getty Information Institute's "aka" project)[14].

Although some improvements can be achieved through such enhancements within the search architecture, it remains questionable whether effective access from a variety of perspectives can be achieved without conscious investment on the part of content providers to identify the nature of their content at a relatively fine level of granularity. The issue becomes particularly acute with the addition of non-textual content (still and moving images; audio materials) given the current lack of effective equivalents to "full-text searching" for these elements. The requirement to develop textual surrogates to identify the content of non-textual materials may, in the short-term, push the development and adoption of sophisticated metadata more rapidly than would occur for textual content.

[13] David Bearman has explored this issue in "Standards for Networked Cultural Heritage", *Archives and Museum Informatics* 9:3 (1995): 279-307.

[14] This application is available on the Getty Information Institute's Web site. <http://www.gii.getty.edu/aka>

In the brainstorming sessions CHIN held with museum directors, participants identified partnerships with related institutions such as libraries and archives as an important opportunity in the provision of information for public access. In principle, audiences could thereby explore, in addition to the objects themselves, a wealth of information about the historical or social context for the objects or local information related to their significance within a given community's history. A key aspect of realizing benefit from such partnerships, however, will be the ability to navigate among the diverse types of information held by the various institutions, in terms of both intellectual access (identification of relevant material) and retrieval of diverse formats.

CIMI has taken some early steps towards realizing this vision in its CHIO project, in mapping between the CIMI DTD and the MARC standard. Further CIMI work in testing the CIMI DTD against DTDs developed within related disciplines such as archives and the humanities, is scheduled for the future. In providing integrated access to content from various disciplines, particularly as we move from referential material (e.g., bibliographic citations) to more substantive information, we will inevitably find that changes will need to be made to metadata approaches developed in an object-centric context. A particular challenge, therefore, will be to find approaches which do not penalize early adopters.

Opportunities for Collaboration

Many of the necessary developments in providing more effective access will take place outside the cultural heritage community. We will need to be aware of these developments and to determine how they can be adopted by, or adapted to, this community. In preparation for emerging opportunities to generate revenues or fulfil missions more effectively, however, the community will also need to collaborate in moving towards the conditions which will allow value-added access to the particular knowledge its members hold. The areas for collaborative action outlined below seem promising.

Peer Review of Sources: In the short-term, heritage institutions would benefit from the identification of reliable sources of information for our own use in managing and researching collections and making them available. As we ourselves make more knowledge available to the public in electronic form, there will be opportunities to enrich the audience experience by providing dynamic links to related knowledge, whether presented by other heritage institutions or made available by other sources. Access to community-level information on the reliability or focus of other sources could reduce the research investment required to achieve this on an individual basis. In the longer-term, the heritage community may want to consider some form of accreditation process for recognized Web sites (or their equivalent in the evolving environment) so that audiences may be assured they have reached "the real thing".

Audience Research: Audience definitions of added value will be contextual, conditioned by the broader information environment. The ability of the community to meet these audience expectations may determine whether museums and other cultural heritage institutions will be active participants in a networked educational and leisure environment. There needs to be a fuller exploration of what will constitute effective access for various audiences. The work which has been launched by the Getty Information Institute on "points of view"[15] is a promising start but it needs to be extended to incorporate not only an understanding of the types of information audiences may be

[15] For an early report on this project, see Sledge, J. and M. Case, "Looking for Mr. Rococco: Getty Art History Information Point-of-View Workshop", *Archives and Museum Informatics* 9:1 (1995): 124-129. CIMI based elements of its DTD on the results of this analysis.

seeking but also how they wish to navigate, how results need to be presented to be effective and what value-added services they would find useful.

Interface Design and Functionality: A key challenge is the need to mediate between and among diverse audiences and content holdings. The interfaces currently provided to audiences are extremely primitive and development needs to be guided by a more sophisticated understanding of audience needs, of the ways in which audiences wish to interact with various types of content, and of the ergonomics of various interfaces. Much of the work in this area will take place outside the cultural heritage community and may ultimately result in approaches in which an individual user can highly personalize his or her interface at the desktop,[16] but specific utilities may require community development. Thesauri are emerging as potentially important in the access process and community efforts to identify additional resources of this nature (or to fill significant gaps) will be essential. Of particular interest, for example, will be the extent to which the existing resources serve the needs of younger audiences. In the global environment, thesauri or other utilities to facilitate access to content in multiple languages may be critical requirements. Investment issues will surface in the time and cost to develop such specialized resources if significant benefit cannot be demonstrated.

Metadata Architecture: Efforts to develop measures which will enable more precise definition of relevant content through metadata are at early stages, particularly when we move beyond text to images, audio and video resources. Collaboration within the community to identify acceptable practices and test the impact will be essential. Testing results against approaches developed by other communities and mapping among diverse knowledge representation models will be equally important. Work to develop the "Dublin core" metadata definition, followed by the "Warwick Framework" and subsequently by efforts to define metadata for images at the CNI/OCLC "Dublin Image Workshop" have advanced the substance of appropriate architectures and provided opportunities for the heritage community to ensure that it is not developing approaches in isolation but the work is far from complete. It is clear that approaches such as marking up content will only be sustainable if the tools exist to minimize the required investment on the part of institutions and/or if the return on investment can be effectively demonstrated. Community efforts to influence vendors with respect to the development of appropriate systems and to share the results of experience may result in a more rapid development of cost-effective solutions, particularly if the efforts can be shared with other communities.[17]

Efficient Intellectual Property Regimes: We have yet to develop an economic model for cultural heritage content in digital form which will both sustain continued development and allow for broad public access. There is strong evidence that cultural heritage institutions managing intellectual property transactions on an individual basis will not realize an adequate return on their investment. (Equally, the cost of identifying and transacting items individually will be prohibitive for value-added producers.) An issue in moving from the single institution model to group models (agents, brokers, sub-licensors or collectives), however, is that there is little evidence on which to base expectations of increased revenue. Projects such as the Museum Educational Site Licensing initiative, supported by the Getty

[16] If the sophisticated, personalized interface does emerge, it may create additional pressure to adopt "open" approaches to content preparation so that content is amenable to varying treatments by desktop clients.

[17] Influencing vendors is a key element of CIMI's work plan.

Information Institute, the Mellon Foundation and the participating institutions, may yield some further insight, as will the practical experience of the consortium being launched by the American Association of Art Museum Directors. Other studies are necessary to address the implications of additional models for the management of rights. Community discussion, both internally and with potential markets, will be essential to develop an understanding of appropriate terms and conditions for use of cultural heritage content.[18]

Business Case for Information Management Strategies: Paul Evan Peters, founding Director of the Coalition for Networked Information, pointed out that we need to move from experiments which provide "proof of technology concepts" to ones which address "proof of program strategy" if we are to succeed in providing our institutions with an appropriate basis for decisions on organizational strategies in fulfilling core missions by becoming electronic information providers.[19] In a 1995 conference on the economics of information in the networked environment, Dr. Hal Varian observed that the "good side of information is that you can sell the same product over and over again. You just sell it in a little different form. When you look at the companies who are going to be successful in this business, they are going to be people who manage to sell the same sort of information in a variety of formats to a variety of audiences at a variety of prices".[20]

[18] Both CHIN and the American Association of Museums have studies underway in this area.

[19] Paul Evan Peters, correspondence with the author.

[20] Varian, Dr. H. "The Economics of the Internet and Academia", in M.A. Butler and B.R. Kingma (eds.), *The Economics of Information in the Networked Environment* (Association of Research Libraries, 1996).

While leaving aside for the moment the question of whether museums will choose to pursue a commercial model, Dr. Varian's observation contains a key concept: organizational strategies for information providers must allow for the continual re-use and repackaging of information. The cultural heritage community will need to develop and test sustainable models which will allow us to realize the benefit of one of our most important strategic assets: the knowledge we have developed in documenting and interpreting our collections.

Conclusion

Generic search engines will continue to play a role for those who wish simply to wander the library shelves in search of something interesting. They are becoming more sophisticated and inevitably will incorporate new capacities which can, as yet, barely be imagined. In a world of extraordinary heterogeneity of approaches to the development of content, however, it seems likely that audiences will be increasingly inundated with non-relevant results. It also remains to be seen whether generic search engines will begin to seek a greater or more direct return on investment than is currently obtained from advertising revenues or licensing arrangements by seeking to generate some form of payment for use over and above the current time-cost to users.

Specialised communities of interest can be expected to seek services which offer more focussed results, interfaces appropriate to the particular audience, greater assurance of the reliability of sources, community communication and the possibility of conducting associated transactions such as rights clearance. Premium service providers are beginning to emerge in the assumption that these communities of interest will be willing to pay for these value-added features. Clearly, sectors which already function on a commercial basis will be better positioned to assign value to the services and, by providing appropriate return on

providers' investments, to foster the conditions which will lead to continued service enhancement.

The cultural heritage community may find itself in a challenging position in such an environment. Services destined only for internal use by the community are unlikely to be commercially sustainable, while services which we would like to develop for a more general audience may need to build acceptance for a commercial orientation. If cultural institutions choose to offer services through commercially-based providers, we will be faced with a philosophical debate revolving around the appropriate equilibrium between core missions and the need for revenue. If, on the other hand, the financial basis for premium service provision in this domain is sponsorship by governments or philanthropic organizations/activities, the community may be faced with compromises in terms of stability of service or service quality (if sponsorship provides immunity from market realities which would drive service enhancement). The community may need to explore options which could provide cross-subsidization between services designed to meet specific market needs (e.g., those of value-added producers) where commercial prices could be charged and "core mission" services for broader audiences at lower or no cost.

There seems little question that institutions will need to invest in managing their information if they wish to exploit the potential of the new media. Institutions which seek to act independently in the provision of information will face large investments in providing an appealing user-environment and may forego revenue or mission-related opportunities which are dependent on easy access by users to large volumes of material and associated transactional services. The cost of participating in multi-institutional or intermediate service arrangements, on the other hand, may be a higher investment in conforming to community standards such as metadata or systems capable of supporting specific exchange protocols.

The economic evidence that could guide decision-making in this area is only beginning to emerge. The economic model will need to take account not only of potential markets and revenues in the networked environment but the cost to institutions of fulfilling their core missions if approaches are adopted which do not facilitate the re-use of content for diverse purposes or the integration of specific content elements with information from other sources. At this point, it is reasonable to assume that the cost of technology and systems to support the preparation and management of an institution's diverse information holdings will decline while the cost of the intellectual effort required to recreate this knowledge for successive uses will rise.

Providing our audiences with effective access to the knowledge we have developed and the stories we have to tell is a core mission which has not changed. What we do need to continue to explore are the implications for our organizations of new ways of delivering this knowledge.

ARTICLE

"Spiders across the stars"[1]—A Web-based Model for Providing Access to Multi-institutional Museum Information

Deirdre C. Stam
Syracuse University

Abstract

Overcoming many of the past barriers to the distribution of museum data, the world wide web has the potential to serve as an effective mechanism for linking the web sites of museums and allied projects. For maximum utility, one would need to make modest additions to the usual discovery/retrieval methods now common on the web. Between the search tools now existing in the wider web context and the many web sites maintained by museums and allied entities, one could interpose expert systems already existing but designed for other purposes, and archival-type descriptions of collections/institutions/projects against which these enhanced web retrieval/discovery tools could operate. The result would be a system that would initially point the user to web sites where his/her inquiry would very probably be successful. The later stages of the inquiry would take the user into local systems where the rules developed for the subject domain in question could be applied through interactive inquiry, thus allowing to the user to refine the question in light of his/her better knowledge of how information in that source is represented.

Background

For over thirty years, at least since the significant IBM conference at the Metropolitan Museum of Art in 1968[2] addressing museum "automation," museums have been looking for a mechanism for automating their data and making it more widely available. There have been variations on this dream, of course. Some institutions wanted only to make information on collections available to in-house staff, some have aimed toward providing archival and collection information to researchers and other museum professionals as well as their own staffs, and some have dreamt of these goals and the wider one of making information about museum riches understood by a wide and general public. The constant in these visions is a desire to publish, in some sense, information about collections.

Progress toward these variations on the goal of wider access to museum data has been erratic and thus far unimpressive, especially when compared to the allied communities represented by libraries and archives. The reasons for slow and uneven progress toward this goal can be grouped under four sets of issues: professional, economic, technical, and administrative. Some of the inhibiting factors in each category are merely circumstantial; there is nothing that museums could have done to overcome these inhibitors. Others, however, represent choices made in the museum profession and in the field with which it enjoys a symbiotic relationship, museum informatics.

Past Inhibitors to Wide Distribution of Museum in Electronic Form

The foregoing factors that have inhibited progress toward general sharing of museum data make up a dreary catalog. They are worth examining, nonetheless, for they are the reality of the museum world. One cannot—and it took some of us 20 years to figure this out—build a model for communicating museum information that does not take into account the realities of the museum subculture and circumstance.

Among the inhibiting factors are professional issues: proprietary attitudes toward museum information; insecurity

1 Kerouac, J., *On the Road* (1957).

2 Metropolitan Museum of Art, *Computers and Their Potential Applications in Museums* (New York: Arno Press, 1968), *passim*.

about the accuracy of institutional records; desire to retain control; fear of public scrutiny of inner workings of the museum; fear of sensationalist publicity from politically motivated "yahoos;" discretion about personal information; and a tradition of privacy and secrecy associated with American cultural institutions derived in part from their elitist roots and past histories.

Economic factors also discouraged wider distribution of museum information in electronic form: insufficient funding for capital investments in computer infrastructure and equipment, unpredictable funding for ongoing costs of personnel and maintenance, and unfamiliarity with budgeting and procurement strategies in this area. It is almost embarrassing to mention, but significant enough to need noting, that many smaller museums have been the recipients of numerous Trojan horses in the form of gifts of outdated and inappropriate computer equipment, given with tax benefits in mind by "friends" and trustees; these gifts have so frustrated staff that they have caused some to reject automation altogether.

Technical issues, too, proved to be barriers to timely adoption of information technology: technology-resistant structures; lack of inhouse expertise; resistance to new technologies on the part of largely humanistically-strained staff and volunteer ranks; the primitive nature of early technologies which were prematurely applied to the complex and highly visual needs of museums; and a discouraging history of early adoption of information technologies under the direction of less-than-qualified and largely self-styled experts—and not a few snake-oil salesmen of both genders—who operated as loners and tended to encourage museums to do likewise. Lack of standards on various levels, and a general resistance to cooperative action in this area, contributed as well to a situation of autonomy and amateurism in computerization efforts.

And finally, administrative issues delayed adoption of information technologies in this arena: unfamiliarity with advances in allied fields; inadequate understanding on the part of old-style administrators and trustees about the nature of modern museum work and use of the museum information base for management purposes; disdain for "techies" and technology implicit in the most prestigious museum training programs; preoccupation with institutional survival; lack of skill in managing technology; a tradition of independence among museums (especially in the United States); and the lingering existence of the old-style trustee whose understanding of the museum as institution was based on circumstances of an earlier era.

Why This Model Will Work

The model proposed here takes into account all of these factors, and to a greater extent than any previous approach, minimizes or bypasses the circumstances and tendencies that have inhibited general publishing of museum data. This model does not require cooperation among participating web sites; it recognizes that museums have not been compelled in the past by this *modus operandi*. It uses a largely *a posteriori*, rather than an *a priori* approach. It assumes that the web interface is becoming the *de facto* distribution mode for museum information, as it is for much other material in similar realms. The model uses the technology of the world wide web, combining standard web search tools and structures with some discipline-specific expert systems, to overcome the disparity among systems to the extent necessary to bring about meaningful retrieval/discovery.

This approach to model-building—the borrowing and adapting of existing technologies with good prospect of continued support and development—gives the product an inclusive character, rather than a competitive nature. Almost all current museum information projects—those from the Museum Computer Network (CIMI, for example), initiatives of the J. Paul

Getty Trust, various rights ventures, and even commercial systems—have a place in this model. The model presented here can be thought of as specifying relationships among evolving tools, and not as a purely technical venture reflecting a particular phase of the automation craft. It thus is capable of evolving along with the internet itself as the demands for computerized museum information mature and expand.

This approach does not actually require any additional work, beyond mounting its web site for its own purposes, on the part of the institution or project included. In fact one can include a museum web site without asking or even telling the sponsoring institution, according to present web mores and practices. One simply plugs in the URL (internet address) of the museum into the model. This approach does not interfere one iota with the current approach to web accessibility that is embraced by the participant. Thus the model is totally unobtrusive for the included institution and the local institution or project retains complete control over the its own system. If the institution/project wishes to be described in an "archival-type" statement against which the multi-institutional search mechanism can retrieves, then that entity will want to cooperate at least to the extent of providing a little structured prose. Required cooperation ranges from nothing to minimal effort. That seems a vital point in a field where cooperation has never been a compelling principle.

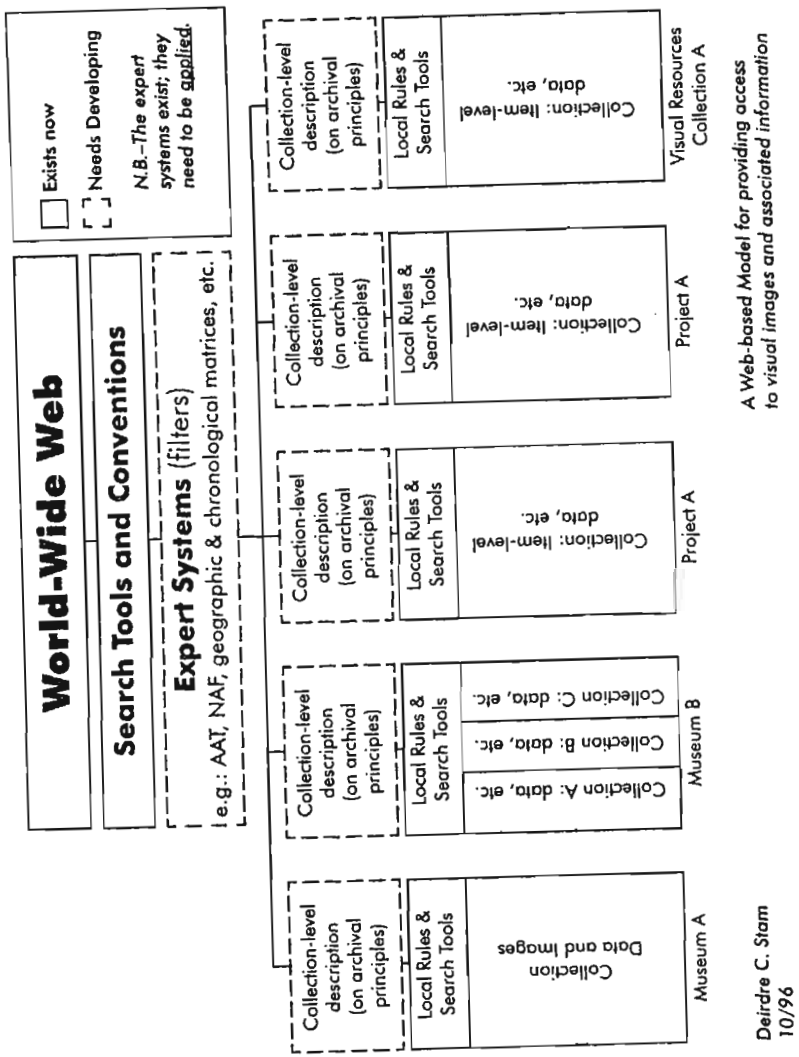
The Kinds of Information Embraced by the Model

The kind of information that was thought to be appropriate for electronic distribution began historically with collection records in systems designed for internal control. It was to gain some internal control over collection records that much of this work in automation began at all. These records were almost entirely textual, with the notable and ingenious exception of early work in Canada[3] that involved adding microfiche images to cards produced through batch processing. While registrars wrestled with text, departments of education went a separate way using videotape, videodisk, and more recently digital imaging technology for products that interpreted the objects and objectives of museum exhibitions and collections. Many of these products were turned into publications, or cannibalized for public relations activities. And also going their own way were photo archives and slide collections which attempted to catalog their holdings and produce images with potential for wider distribution, and longer and more robust life, than their then current technologies allowed. Interestingly enough, it is only with web technologies that one is seeing these streams of museum information activity converge, and that convergence is uneven due largely to administrative separation of the origins of these kinds of information even within one institution. The model presented here assumes that these streams of information will remain largely separated due in part to their different functions within the museum context, but that the inquirer would in most cases like access to the totality of information from an institution as a unit. It is assumed here that content, and not the administrative structure of the source, is of interest to most web inquirers.

Parts of the Model Exist Already

Many of the parts of this model exist already, and they are already publicly accessible. Beginning from the top of the schematic, these elements bounded by solid lines already exist: the web, searching tools and conventions, systems and local search mechanisms, and information sets from specific institutions consisting variously of collection data, catalogues

[3] Sledge, J. and B. Comstock, "The Canadian Heritage Information Network", in R.B. Light, D.A. Roberts and J.D. Stewart (eds.), *Museum Documentation Systems: Developments and Applications* (London: Butterworths, 1986), 7-16.



A Web-based Model for providing access to visual images and associated information

raisonnes, interactive learning materials, operational information (such as hours of opening and exhibition schedules), and allied project materials. While this model specifies museum data, one can image adding material from visual collections, archival data, and other web-based data sets as well. The number of web pages containing such information is proliferating at an astonishing rate.

Also existing, but not applied yet to such an approach, are expert systems. These are structured knowledge bases. Some of these were developed for other purposes, such as cataloguing specimens or indexing catalog data. The latter include terminology sets for object types and art/architectural forms. Other examples are tables of personal names and variants, geographic information, and matrices for chronological and period information. Some of these expert systems are publicly available, but many are currently associated with single projects or institutions, and are being recreated in part by other projects and institutions attempting to make sense of similar data sets. Why this should be so, and how such insularity might be overcome, is a subject well worth exploring at another time.

Existing in fragmentary form are the necessary prose descriptions of project and institutional collections (called "collection-level descriptions" in the diagram) that indicate, much as an archival record would, where the collection came from, what it includes, how it is structured, what its chronological and geographical limits are, what intellectual or historical principles govern its development, and how it has changed over time. One could use descriptions from standard museum directories, respecting copyright of course, but some richer and more consistent descriptions would make the entire concept more effective. Strategically, one could begin with information already existing in web site introductions or standard museum descriptions (and mark it up appropriately with SGML) and then in time, when museums are convinced that it is worth the time to cooperate, ask

for a few pages of structured data for this purpose from projects or institutions themselves.

The Nature of Retrieval in the Model

The kind of retrieval supported at the upper, internet level might be called "discovery." That is, the inquirer poses his or her question, and that question is translated by means of search algorithms and translation processes through expert systems into terms that are meaningful at the individual collection-institution-project level. The inquirer is taken to the general "neighborhood" where he or she might find an appropriate response to the question posed. At that point, the inquirer enters the local systems, each of which has its own conventions, rules, and systems features. He or she must then pursue the inquiry within the rules of the local web site. The approach thus assumes that browsing will be a fundamental need in the discovery process, in that the question can be refined interactively and dynamically as the inquirer finds out in subsequent steps how his/her topic is discussed within the confines of specific (intellectual and technical) domains. It is assumed that the inquirer wishes to negotiate his/her own inquiry, and requires mainly of a system that it point to sources where the likelihood of success in refining the question and finding suitable answers to it will be great.

One finds pieces of this approach in some web retrieval tools, such as Yahoo, where one can specify a general subject area, and then attempt a word-match request within that domain. The approach advocated here adds several features. It translates the language of the inquiry into the terminology of a fairly narrow intellectual domain, and then offers not specific pages that carry the term specified as happens with most web tools, but rather projects or institutional information data sets, as a response. The vital question of authority is thus addressed—the inquirer is very aware of the institutions with which he/she is negotiating the query—and the inquirer is able to control the later stages of the search as he/she understands better, in response to interactivity with a data set, what the range of possible answers might be. The structure he/she encounters in the later stages of the search has been established by the institution or project itself, and presumably represents the best thinking of experts on how one reasonably and fruitfully negotiates a question within the subject area under consideration. Do you want to see a frog? Let us—The Official Frog Museum of North America—show you how we think about frogs, and offer you some options as to the kinds of frogs you can find. You can then use our system effectively to find the frogs that best meet your newly defined criteria.

How Can the System be Developed?

There remains the very serious question of how this system can be developed and where it might be maintained. A look at the past would indicate that attempts to set up stand-alone organizations to do systems work in the museum world stand little chance of succeeding. Why this is so is not fully understood, but the fact remains that such an approach seems not to work. Excellent work on content issues and related expert systems can and does take place in professional societies and research institutes. System work, however, requires an infrastructure of equipment and expertise that seems to exist, at this time, mainly in universities. That would seem the best locale for such development work. In such places, work of this kind is commonplace. The amount of development work necessary to tie together these pieces to create a test environment is quite modest. This is especially true if every effort is made to employ tools and methods that have currency in the evolving internet.

The Future: Administration, Implementation, and Maintenance

The question of future maintenance is not easily answered from looking at past practice since the past did not have the

multitude of business models, and players, and institutions with which to interact. The answer to this question may depend in part on what the test environments yields. Should the results break new ground in retrieval, then the commercial world may want to adopt the work as part of larger tools for retrieval. Should the results greatly improve work in a particular domain, such as the humanities, a research institute in this area might want to take on the task. Should the text results indicate that a distributed mode could be adopted, then the museum world may want to look closely at models for distributed work being explored by the research library community and other allied organizations dedicated to scholarly computing. The point to keep in mind is that the adoption of open standards and evolving tools used for the internet generally will keep the costs and complexities of maintaining a service of the kind described here at a minimum.

The proposed model represents a way of thinking about tying together various components that could work together for better retrieval/discovery of museum-based information on the web. The existence of such an effective method could well inspire other institutions and projects to hasten their own web interfaces, thus leading to rapid expansion of the content base for the concept. The model is certainly not revolutionary. That is really its strength. It is based on much that is known or understood by analogy already. Its virtue is the ability to relate parts now seen as separate and to make sense of a technological and administrative environment, online museum information, that is daily growing in richness and complexity.

CONTRIBUTORS

Wendy M. Duff is an Assistant Professor at the University of Toronto Faculty of Information Studies. She received her Ph.D. from the University of Pittsburgh's School of Library and Information Studies. Her primary research interests are electronic records and archival description. She was the project manager for the University of Pittsburgh's electronic records project which developed the Functional Requirements for Electronic Evidence. She presently teaches courses on record management, electronic records management, and archival description.

Margaret Hedstrom is an Associate Professor at the School of Information, University of Michigan, where she teaches in the areas of archives, electronic records management, and digital preservation. Before joining the faculty at Michigan, she worked as Chief of State Records Advisory Services and Director of the Center for Electronic Records at the New York State Archives and Records Administration. Dr. Hedstrom earned master's degrees in Library Science and History, and a Ph.D. in History from the University of Wisconsin-Madison. She has published widely on many aspects of archival management, electronic records, and preservation in digital environments.

Lyn Elliott Sherwood is the Director General of the Canadian Heritage Information Network.

Deirdre C. Stam is currently coordinator for web activity at Syracuse University and is an adjunct faculty member the School of Information Studies at SU. Formerly executive director of the Museum Computer Network and curator in several art museums, Dr. Stam has taught library and information science at Columbia University and Catholic University, and has published widely on issues of computer applications to museum and art information.

