Preserving Digital Archives in the 21st Century

Wendy Duff

Two months ago, í was visiting a friend and found myself watching her four year oíd play on the family computer. As I drew near to see what kind of game he was playing, I discovered that he was reading a Winnie-the-Pooh story. Unlike the Winnie-the-Pooh books that I had read to my chiidren, this story was interactive. The old black and white iliustrations by Ernest Shepherd had been replaced by fiiil colour and talking characters. My young friend Alex, could make Pooh move, and occasionaliy the bear on the screen would say," oh dear " Alex was very excited, and he seemed to gain great delight from ínteracting with this digital book. As I watched him I thought about the difTerences between his digital book and my paper-based version of Winnie-the-Pooh. The digital book was dynamic, and interactive while my paper-based book was linear, and static. The digital book was bright and giitzy while my paper-based book was rather plain and simple. However, the digital book needed compatible hardware and software to work, while my paper-based book could be accessed by anyone who could read. Furthermore, the digitai book would have a very short life-span of five or so years because of hardware and software obsolescence, while my paper-based version had been read to my chiidren and may one day be read to my grandchildren. This interactive, coiorful, giitzy digital version of Winnie-the-Pooh reminded me of the many challenges archivísts and librarians face when they try to preserve this new médium.

The management and preservation of electronic records present complex and challenging issues to information professionals. The gravity of these issues led the National Association of Government Archives and Records Administrators* (NAGARA) Advanced Institute for

Government Archivists to conclude in 1990 that "archival management of electronic records is probably the most important, and certainly the most complicated, issue currently before the archival profession."[1] Eight years later archivists are still challenged by digital archives because: 1) electronic records are system dependent, 2) electronic records exist on fragile media; and 3) electronic records can be easily erased or changed.[2] Just like my young fhend's version of Winnie-the-Pooh, our digital heritage will become unreadable unless organizations design systems that ensure its long term preservation.

Moreover, the preservation of electronic records presents even greater challenges than the preservation of books ín a digital form. Maintaining the hardware, software and/or médium does not ensure that records will continue to serve as adequate evidence of actions and events. John McDonald suggests that an electronic record can become inaccessible if:

- it lacks sufficient documentation to permit ongoing intellectual understanding of its content and context
- it access is dependent on software and hardware that can be expected to change over time
- accountability has not been assigned for ensuring that such information is identifíed and protected.[3]

Moreover, for a record to serve as evidence, it must be reliable, authentic and understandable. Duranti points out that

> [rjeliability is provided to a record by its form and procedure of creation .A record is

National Associáation of Government Archives and Records ***AámimstTaioÉs-fhivalAdmimstration in the*** *Electronic Age: An Advanced ínstiiute* ***for Government Archivists,*** **[co-sponsored by the School of Library and Information Science. University of Pittsburgh, and funded by the Council on Library Resources] (Pittsburgh: NAGARA,** 1990). 2.

[2]United States. National Historical Publications and Records Commission, *Electronic Records: A Report to the Commission,* Corrimission Reports and Papers 4 (Washington, D.C.: the Commission, 1990)

[3]John McDonald, "Managing Information in an Office Systems Environment: The IMOSA Project,"[1] *The American Archivist* 58 (Spring 1995): 143.

> considered reliable when its form is complete, that is, when it possesses ali the elements that are required by the socio-juridical system in which the record was created.... The procedure of creation of a record is the body of rules according to which acts or portions of them are recorded. Some of these rules refer to records-makers. Other rules refer to the routing of the records, their filings,[etc] ... [Authenticity] warrants that the record does not result from any manipulation, substitution, or falsification occurring after the completion of its procedure of creation.[4]

Therefore to preserve an electronic record, we must ensure that the record can be read by current hardware and software facilities, and we must also ensure that the record was created and maintained according to proper procedures and controls. We must have information about the business process from which the record emanated. We need to know who created the record and whether they have the required competencies to act as they did. Furthermore, we need to know whether the system that maintained the record was a trusted system. Society will not trust a record unless they know where the record carne from, how it came into being, what it documents and how it was maintained. Capturing and preserving this information along with the record is essential if one is to preserve its evidential value.

Over the last five years, a number of archivists and records managers have endeavored to identify strategies that meet the challenges of preserving digital archives. These projects are quite diverse. A few initiatives have examined governmental policies and practices to identify both obstacles that hinder, and best practices that support, good electronic records management.[5] Other research has attempted to identify the elements, the requirements, and the environment

[4] Luciana Duranti, "Reliability and Authenticity; The Concepts and Their Implications," *Archivaria* 39 (Spring 1995): 6-8.

[5] **For example see New Yoris State Archives and Records Administration. Center for Electronic *Rsco&silding Partnerships for Electronic Recordkeeping: Final Report and Worláng Papers* (Albany: New York State Archives and Records Administration, Center for Electronic Records, 1995); T.K. Bikson and E.J. Frinkfikgpe/v/ng *the Present: Toward Viable Electronic Records* (The Hague: Sdu Publishers. 1993). International Council on Archives. Committee on Electronic Records(juúfe *for Managing Electronic Records from an Archival Perspective* (Ottawa: International Council on Archives, 1996).**

needed to create and preserve reliable and authentic records.[6] From these various studies agreement has emerged concerning the fundamental nature of records,[7] the components of a record,[8] and the importance of knowing the regulatory and legislative environments in which records are created.

Time does not enable me to speak in detail on ali these projects but I would like to use the time remaining to discuss the fmdings of research conducted at the University of British Columbia (U.B.C) and the University of Pittsburgh (Pitt Project) . I have chosen these two projects because of the subject of their research and because of their significance. Finally, I will conclude with some remarks on their implications for information professionals.

The U.B.C. Project

The UBC project was a two-year research study to establish, in principie, what a record is and how to recognize it in an electronic environment.[9] The focus of the project was current

Luciana Duranti and Terry Eastwood, **"Protectíng Electronic Evidence: A Progress Report on a Research Study** and Methodology/>ÍA-c/7/v/ ***& Computers*** **5 (Fase 3, 1995): 213-250. Luciana Duranti, Heather McNeil and William E.** Underwood, **"Protecting Electronic** Evidence: **A Second Progress Report on a Research Study and Methodology,"** *Archivi d- Computers* **6** (Fase **1, 1996): 37-69** or see **their web** ***siiéittp://www.slais.ubc.ca/users/duranti/;*** Or **see** *http://www. sis.pilt. edu/~nhprc/.*

Although the definitions of records **differ slightly, ali of them support the proposition that records are evidence that** arise out of the conduct of business activities or **transactions. For example the ICA guidelines defines records as:** 'recorded information produced or received **in** the **initiation, conduct, or completion of an institutional or individual** activity and that comprises content. **context, and structure sufficient to provide evidence of the** ***2&ú\fffl$i4e for*** *Managing Electronic Records From an Archival* ***Perspective,*** **13), while the UBC project defines records as:** "documents created by a physical or **juridical person in the course of practical activity"** *(http:, 'www.siais.ubc.ca/users/durantij.* The **Building Partnership Project defined records as: "ali books, papers,** microfonns, computer-readable **materiais, maps, photographs, film, video and sound records, or other documentary** materiais, regardíess **of** physical **form or characteristics made or received by any agency or by the legislature or the** judiciary in pursuance of law or in **connection with the transaction of public business and preserved by that agency or its legitimatesuccessors** as evidence **of** the **organization functions, policies, decisions, procedures, operations, or other** activities, or because **of** the **information contained thereinfJv'** ***Iding Partnerships,*** **6.**

[g]

Although there is some variation in **the terminology used by the projects noted above, there is agreement that** records require content, structure **and context. The ICA guidelines define records as consisting of content, context and** structure, while the **Pittsburgh Project funcuonal requirements emphasize that records must be aceurate, understandable,** and **meaningful** which **relates to capturing the content, structure and context. The UBC project states that the "requiremenls for** any record **to** be **created (made or received) are: a. médium; b. content (faets or information); c. form;** d. person (author. writer, addressee **and creator);** e. **acts; f. archival** ***bonUltp./Avww.slais.ubc.ca/users/duranti.***

records rather than archival documents. The project's methodology was based on the assumption that solutions to electronic records problems cannot derive from purely pragmatic approaches. Solutions must be based on principies and concepts that are applicable in many different situations. The project used the deductive method of research and grounded the research on the principies of diplomatics. The project staff teamed up with the United States Department of Defense Records Management Task Force and used the techniques of IDEF modeling. IDEF data modeling analyzes activities to identify the inputs, and outputs of the activities, the entities and principies that control them, and the mechanisms needed to carry them out. The project identified the juridical system, the creator's mandate and fiinctions, national and international standards, and the principies of archival science as the overarching elements that control the management of an organization's records. They used IDEF modeling to decompose each activity into its constituents parts. Working with the US Department of Defense, adhering to diplomatic principies and concepts, and using IDEF modeling, the project was able to delineate the activities needed to manage records and develop detail procedures for managing records in the electronic recordkeeping systems. The project concluded that the reliability and authenticity of electronic records:

- are best ensured by embedding procedural rules in the overall records system and by integrating business and documentary procedures;
- are best guaranteed by emphasizing their context; and
- can only be preserved if they are managed together with ali the other records beionging in

Luciana Duranti and Terry Eastwood, "Protecting Electronic Evidence: A Progress Report on a Research Study and Methodology," *Archivi & Computers* 5 {Fase 3, 1995} . 213-250

the same fonds.

Furthermore, it suggested that a complete record would need the following components. [11]

a. médium. A record needs to be captured on a médium.

b. content. A record must contain facts or information

c. form. A record needs to have a readable and intelligible form.

d. persons. A record must have an author, a writer or originator, an addressee, and a creator.

e. acts. A record needs to be directly connected with some action.

f archival bond. A record needs to be part of the whole of the documents made or received in the course of the activities of its creator.

The results of this research provides interesting insights especially when they are studied in conjunction with the findings of the Pittsburgh Project.

The Pittsburgh Project

The University of Pittsburgh Project was a three year study funded by the American National Historical Publications and Records Commission to investigate electronic records issues.

The project began by assembling an interdisciplinary team of library, information, and archival science experts to delineate the functional requirements for recordkeeping. The team identified 13 functional requirements for capturing, maintaining and using reliable records.

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

- **requirements that relate to the organization - labelled Compliant**

[10] See *http:, www. s/ais. uhc. ca/users/duranti/*

[13] *http:•/•"www. slais.uhc.ca/users/duranti/*

organization;

- requirements reflecting specifications for recordkeeping systems - classified as Accountable recordkeeping systems;
- requirements that relate to the record - these are grouped under three sub-categories: Records - Captured; Records-Maintained , and Records-Usable.[12]

The first requirement, Compiiant, highlights the importance of knowing the legal and regulatory environment in which an organization functions. This environment will vary from country to country and from industry to industry. Companies such as the pharmaceutical industry are highly regulated while other types of businesses have fewer explicit laws to guide their recordkeeping practices. This requirement confirms that the juridical system, standards and best practices dictate specifications for keeping an organization's records.

The next group of records relate to the recordkeeping system. The environment in which records reside can either increase or decrease their reliability and trustworthiness. The admissibility of records depends upon testimony that verifies the integrity of the recordkeeping system that controlled them. Therefore, the second group of requirements delineate specifications for the recordkeeping system.

The third category of requirements specify characteristics that records

[12] For a fiill description of the functional requirements and a history of the project see *http ://www. sis.pitt. edu/~nhprc*

must have. These requirements specify that a system must capture, maintain and make usable the content, structure and context of records. Furthermore, records must be authorized, e.g., only people with appropriate authority can create records. Records must be removable, and parts of records must be able to be masked or redacted.

The project expressed these requirements as production rules. Production rules ensure the specifications are unambiguous, precise as possible, use consistent language, and are measurable. Metadata specifications were derived from the production rules. Metadata are simply data about the records that are needed to ensure that the records are accessible, understandable, and that they maintain their evidential value. The metadata specifications are clustered into six layers: handle, terms and conditions, structure, context, content and history of use. The handle layer contains information needed to retrieve the record. Terms and conditions consists of information needed to access the record, e.g. it documents any restrictions on retrieving or using the record. The structure layer consists of metadata about data structure designed to permit the record to remain evidential over time and to be migrated to new software and hardware dependencies as necessary. The contextual layer identifies the provenance (i.e. the person, system, or instrument that is responsible for generating the record) and provides data that supports the recordas use as evidence of a transaction. The content layer contains the content of the record, and finally the use history layer

documents important uses of the record.[13] The metadata "guarantees that the record will be usable over time, only accessible under the terms and conditions established by the creator, and have the properties required to be fully trustworthy for purposes of executing business."[14]

Once the requirements were established, the project staff sought ways to increase the probability that organizations would implement the requirements. The project was aware that the requirements would require modifications to current systems design and some of these changes would fly in the face of information systems thinking. For example, recordkeeping systems require redundant, time-bound information while information systems strive to contain up-to-date data and to reduce data redundancy. Furthermore, archivists and records managers do not have the authority or moral suasion to convince, or force organizations to do it their way. Therefore the team attempted to develop techniques for implementing the requirements that would find resonance in the organizations we needed to influence.

Communication research provided insights into how to influence people. Since the 193 Os, experiments have examined different factors that increase or decrease the acceptance of a message and its ability to influence an audience. Numerous experiments have demonstrated that messages from highly credible sources are more influential than messages from less credible sources.

This research sparked the interest of the project team. It seemed to support the team's hypothesis that archivists could influence the acceptance of the functional requirements by linking a requirement with warrant, that is, statements drawn from the law, standards and a person's authoritative professional literature and best practices. The team argued that "if professionals in

[13] *http://www.sis.pitt.edu/~nhprc*

[14] **David Bearman, "Towards a Reference Model for Business Acceptable Communications,"**

our society were made more aware of the functional requirements for recordkeeping as expressed in recommended practices of their own professions (which are themselves grounded in law), they would be more inclined to take responsibility for the adequacy of their recordkeeping practices."[15]

I decided to test this hypothesis for my dissertation. I wanted to see if I could increase the importance of the requirements by linking them to laws, auditing standards, and professional best practices. I used statements drawn from authoritative legal, auditing and information sources and tested their impact on the evaluations of the functional requirements by lawyers, auditors and information technologists The study found that warrant significantly influenced the rating of importance given to two functional requirements by ali participants. However, it increased rating of importance given by lawyers for four of the requirements. Furthermore, these differences were usually a result of the warrant drawn from the law. The results showed that the warrant, and in particular legal warrant, had its greatest influence on lawyers and that evaluations given by information specialists were not significantly increased by the presence of warrant.

The concept of warrant is fundamental to the recordkeeping profession. We need to understand what other professions and disciplines require of their records and recordkeeping systems. We need to know the requirements promulgated in standards, regulations, etc, so we can advice our organizations on how to keep records. We need to ground our recordkeeping specifications on the foundation of warrrant.

Near the end of the three year study the Pitt Project carne up with a model for ensuring organizations keep reliabíe records. The model postulates that the design of recordkeeping

http:www.sis.pilt.edu/~~nhprc

[15] **David Bearman et al., 'The Warrant for Recordkeeping Requirements,"** ***university of Pittsburgh*** *Recordkeeping Functional Requirements Project: Reports* ***and WorMng Papers, LIS055/LS9400f*** **(Pittsburgh: School**

systems should begin with warrant from which one identifies the requirements for reliable recordkeeping systems. These requirements are then expressed as production rules and finally metadata specifications are derived from the production rules. This model is perhaps more importam than the requirements themselves. It is the lasting legacy of the Pitt project.

Conclusion

In 1991, Lisa Weber noted that "if you ask for a seat at the table, you better have something to say."[16] The U.B.C and the Pitt Projects have provided archivists with an important message to communicateb. Both projects specify the components of a complete record, and identify metadata elements needed to ensure the preservation of authentic and reliable records. They both highlight the need to understand the regulatory environment. U B C posits that records must be managed according to the juridical system, and national and international standards. Pittsburgh argues that functional requirement for recordkeeping and metadata specifications derive from the law, customs, standards and professional best practices accepted by society and codified in the literature of different professions concerned with records and recordkeeping. The U B C project suggests that recordkeeping requirements are universal and should derive from diplomatics and archival principies. The Pitt project argues that requirements are established by society and may differ from country to country or in different cultures. U B C stresses the need for detailed recordkeeping procedures, including the development of classification systems and retention schedules. The U.B.C, project developed a set of rules for enduring the maintenance of reliable and authentic records. On the other hand, the Pitt Project opined that there are numerous

of Library and Information Science, University of Pittsburgh, September 1994), p.[l].

[6] Lisa Weber. "The Working Meeting on Research Issues in Electronic Records: A Report to SAA," unpublished, 28 September 1991.

different tactics for meeting the requirements and an organization's choice of tactic should take into consideration the Corporation's culture, as well as its technological and juridical environment. The Pittsburgh requirements are very high levei and abstract while the U.B.C. Project's rules are specific and concrete. Both projects provide models for the development of requirements for recordkeeing systems.

This research can and should be used by organizations and governments. For example, the Pittsburgh Project provides an excellent model for the identification of recordkeeping requirements. The model requires that each organization compile warrant that pertains to its industry or jurisdiction and dictates specifications for recordkeeping. The requirements specified in the warrant must be identified and delineated as production rules, and metadata specifications. Furthermore, the organization will need to develop tactics for implementing the requirements and ensuring the capture and maintenance of reliable and authentic records.

The Pittsburgh and U.B.C, projects have importance messages, and both realize the need for more research in this area. The findings of these two studies still need to be tested in different organizational settings. However, they have highlighted the importance of understanding the warrant or the juridical system, developing accountable recordkeeping systems that follow recognized procedures to ensure the preservation of our digital heritage.