Provenance, Journal of the Society of Georgia Archivists

Volume 28 | Number 1 Article 3

January 2010

Through to Cyberspace: And What Janus Found There

Richard Pearce-Moses
Clayton State University

Through to Cyberspace: And What Janus Found There*

Richard Pearce-Moses

A few years ago, I was presenting at a workshop in electronic records management for state agencies in Arizona. Many in the crowd came from agencies that had done little or no thinking about how they would manage their electronic records. They had basic questions, like "How long do I need to keep email?"

The attendees were not happy to hear that the messages needed to be filed by content as retention period was based on the content, not the means of delivery. I pointed out that email may be the most challenging problem of electronic records management. Organization is difficult at best because the messages were managed – more usually unmanaged – by the recipient. Likewise, disposition was usually at the users' discretion. Complying with discovery or open records requests was incredibly complicated, because any single message that should have been deleted could still be on any number of desktops, Blackberries, and personal computers at home. Moreover, transferring those messages that need to be kept permanently to the archives was no trivial matter.

The attendees wanted a simple answer, a specific period of time for all email. One fellow commented that managing electronic records should be easier at that point in the information age. The reality is we are not that far into the information age. Ford introduced the Model T in 1908, making cars widely accessible.[1] A hundred years later I rarely look under the hood of my car. I have looked under the hood very rarely in the last fifteen years. However, when I bought my first car in 1974 – a Volkswagen Beetle, which was considered very reliable at the time – I was regularly under the hood. To keep the engine running smoothly, I had to gap the valves on a regular basis, change the points and condenser, and check the timing. That was nearly seventy years after the Model T.

By comparison, dating the origin of the information era with ENIAC (*Electronic Numerical Integrator and Computer*) in the mid 1940s would be like starting with Karl Benz's patent for the automobile in 1886. The IBM 1401 Data Processing System, introduced in 1959, might be a better marker because it was the first system to sell more than 10,000 units.[2] However, the IBM PC, introduced in 1981, might be the most equivalent milestone in terms of popularizing the computer and putting it in the hands of non-technical people. The Apple II and the Kaypro came before the PC, but they were never as pervasive as the PC.

A century after the introduction of the Model T, cars require little maintenance. Given the thirty years since the introduction of the PC, computers are a relatively new

---

[1] "After 20 years of experimentation, Henry Ford finally saw the fruits of his labor in October 1908, with the introduction of the Model T. "The Model T Put the World on Wheels," Ford Motor Company website, http://www.ford.com/about-ford/heritage/vehicles/modelt/672-model-t (checked 10 October 2010).

[2] "1401 Data Processing System," IBM website, http://www-03.ibm.com/ibm/history/interactive/index.html#/FoundationsOfModernComputing/1401DataProcessingSystem (checked 10 October 2010).

technology. It should be no surprise that all the problems are not yet worked out, that IT systems are not as reliable as cars.

A little more than a decade ago, the National Archives and Records Administration first began plans for the Electronic Records Archives (ERA). Ken Thibodeau recounts that in 1998 few archives in the world had experience preserving electronic records, that only the simplest forms of electronic records could be preserved, that those methods were not scalable to the increasing number of electronic records, and that the archival profession had not yet provided a firm theoretical basis for long-term preservation and access.[3]

Some may believe that, from an archivist's perspective, things haven't really changed that much since then. I believe that the records management and archives professions have made significant progress over the past twelve years. At the same time, I believe that there is much work to do and that the work will be hard. I offer some personal thoughts on the state of digital archives.

Archivists are No Longer in Denial

Elizabeth Kubler-Ross observed that when faced with grief – especially with death – people regularly respond in five stages: denial, anger, bargaining, depression, and acceptance.[4] Given the profound impact of digital information on the records professions, records managers and archivists are faced with the death of the old way of doing things.

---

[3] Kenneth Thibodeau, "The Electronic Records Archives Program," Bruce Ambacher, ed., *Thirty Years of Electronic Records* (Scarecrow, 2003), p. 92. Note that the National Archives' work with electronic records predates development of ERA, with the first acquisition on 16 April 1970. See Thomas E. Brown, "History of NARA's Custodial Program for Electronic Records," in *Thirty Years*, p. 1.

[4] *On Death and Dying* (Macmillan, 1969).

When I served as president of the Society of American Archivists (2005 - 2006), I was very concerned that most archivists were stuck in denial. I would often hear colleagues say, half joking, that they would deal with electronic records by retiring. Such a statement was less amusing when made by someone in their twenties. I was somewhat pessimistic about the future of the profession. If records managers and archivists did not step up to the plate, someone else would take their place. Many information technologists did respond, with the result that today many senior executives turn first to their IT shops for advice on electronic records.

Richard Pearce-Moses addresses attendees at the 2010 Society of Georgia Archivists annual meeting

During the year I was president, I worked hard to engage the profession in a discussion about electronic records. I talked to a lot of records professionals about their

response to the digital era.[5] Fortunately, my pessimism was unfounded. I learned that most archivists did not have their heads in the sand, although their attitudes and approaches varied considerably.

Many archivists did not see a great need to learn technical skills. They indicated that they can hire someone with those abilities. I question if this approach is truly viable. Without technical knowledge, how will they know if the solution provided addresses the problem or if it is reasonable and sustainable? More than a few suggested that the next generation of archivists, who grew up with computers, would have the necessary skills. Unfortunately, the skills to use desktop applications, send email, and surf the web are not the skills that archivists need to preserve and provide access to the records.

In 2006, a group of archivists with practical experience working with electronic records came together at the *New Skills for a Digital Era* colloquium to address that question.[6] The attendees noted that archivists need a robust, technical understanding of the very nature of electronic records in terms of media and formats. The participants also saw a need for familiarity with more technical skills, such as database management systems and query languages, markup languages, and file transfer.

One insight that surprised me, though, was a need for "soft" skills. To thrive in the digital era, archivists need to

---

[5] See Richard Pearce-Moses, "President's Message," *Archival Outlook*, September/October 2005 and following issues; http://www2.archivists.org/sites/all/files/AO-SepOct2005.pdf (checked 18 Oct 2010).

[6] See Richard Pearce-Moses and Susan E. David, *New Skills for a Digital Era: A Colloquium Sponsored by the National Archives and Records Administration, the Society of American Archivists, and the Arizona State Library, Archives and Public Records, 31 May – 2 June 2006* (Society of American Archivists, 2008). http://www.archivists.org/publications/proceedings/NewSkillsForADigitalEra.pdf (checked 18 October 2010).

work with a wide range of people. No one person has all the knowledge. Some of the most important skills records professionals can have include communication, negotiation, and facilitation.

Today, I think that a large number of archivists continue to struggle with electronic records because they lack technical skills. Archivists who are willing to get those technical skills are uncertain which ones they need. Fortunately, a number of archival educators have seen the need for formal education. Records professionals can get excellent training through programs at the School of Information and Library Science at the University of North Carolina, Chapel Hill and through the University of Arizona, to name only two.[7] Clayton State University, in Morrow, Georgia, next to the National Archives Southeast Branch, has just started a program that focuses on digital archives.[8]

Through the Looking Glass

Records professionals are much like Alice through the looking glass. As they enter the strange space of electronic records, they see a world transformed. They see things in a new light. At the same time, they see a reflection on what they already know.

What do records professionals need to know to thrive in the digital era? I would answer with a question. What do they need to know about paper records (and other analog formats)? I began programming on a Teletype in 1968, and I began working seriously with the problem of digital

---

[7] See "Concentration in Archives and Records Management," http://sils.unc.edu/programs/arm (checked 30 October 2010) and "Digital Information Management," http://grad.arizona.edu/live/programs/description/272 (checked 30 October 2010).

[8] See "Master of Archival Studies," http://cims.clayton.edu/mas/ (checked 30 October 2010).

archives about ten years ago. Allow me to offer some ideas, based on my own experience and observations.

Most records professionals are unaware of what they know about paper formats. We grew up with paper. Much of our knowledge is tacit and unarticulated. The more time we spend with records in cyberspace, the more aspects of paper records and paper-based recordkeeping systems come alive. Over time, cyberspace becomes less strange and scary as we recognize the familiar in the new. Digital signatures and public key infrastructure? Not too far from chirographs, a technique used for centuries to authenticate records. Luciana Duranti used diplomatics, which has its roots in the seventeenth century, as a starting point to think about electronic records.[9]

Similarly, I think most archivists gain new appreciation for what they know about paper records when they start studying digital information. (By analogy, I really learned English grammar only when I studied German.) In this new context, concepts that were vague or assumed stand out in relief. The underlying archival principles take on new clarity.

Entering cyberspace, archivists begin to learn new terms almost immediately. They can name things that they had never really thought about before. For example, when I worked in historical collections I seldom thought about the authenticity and integrity of the records. Once, I questioned if a description on the back of a photograph was trustworthy. It was an early 20th century photo and the caption on the back did not seem to match the image. The

---

[9] *Diplomatics: New Uses for an Old Science* (Society of American Archivists and the Association of Canadian Archivists in association with Scarecrow Press, 1998). For the source of diplomatics, see Jean Mabillon, *De re diplomatica libri vi. in quibus quidquid ad veterum instrumentorum antiquitatem, materiam, scriptuam, & stilum* (Luteciae Parisiorum, sumtibus viduæ L. Billaine, 1681), citation from the catalog of the Library of Congress, http://lccn.loc.gov/07006236 (checked 18 October 2010).

caption was neither signed nor dated. More significantly, it was written using a felt tip pen, which meant it was significantly after the date the photo was made. While I questioned the caption, I never thought of the discrepancies in terms of authenticity and reliability. Working with digital materials, I understand those concepts much better, and understand why I was troubled by the caption.

Capturing publications from the web requires decisions about how far to follow links. What are the boundaries of the publication? Does a link point to an integral part of the document, or is it external information used as a reference? Include too many links, and the document could include the entire web. With print documents, the question is moot; they have boundaries. "Four-corner" documents have a first and last page, and the pages have limited dimensions. A staple is metadata made tangible, offering information (sequence and contents) about information (the pages themselves). What seems to be a trivial notion in paper has significant implications in cyberspace.

Spending more time with websites, it becomes apparent that many have a lot in common with archival collections. The individual or organization that produced the site is the provenance. The directory structure is analogous to series and subseries. By looking at websites as archival collections, rather than individual publications, it is possible to use archival methods to appraise, acquire, and describe the materials more easily.[10]

Although many things in cyberspace have a certain familiarity, they are not exactly the same. Correspondence and email have clear parallels. In spite of the similarities between paper and digital records, the formats are

[10] See Richard Pearce-Moses and Joanne Kaczmarek. "An Arizona Model for Preservation and Access of Web Documents," *DttP: Documents to the People* 33, no. 1 (Spring 2005). Online at http://members.cox.net/~pearce-moses/Papers/AzModel.pdf (checked 18 October 2010).

sufficiently different that approaches to managing them the same way may fail.

If people received their emails, Tweets, and texts on paper, they would take steps to manage the volume. At some point, their desks would be so cluttered they would have to throw away the useless messages and they would have to file the rest to find them. Otherwise, they would never get in their offices or get anything done. In the realm of paper, records managers took advantage of the filing system to manage retention. In the digital era, space and access are no longer a problem. People resist discarding and filing messages. "Get a larger hard drive! They're cheap!" and "Why file? It's easier to just search my inbox, even when it has 10,000 messages!" Records professionals – especially those involved with discovery and litigation – know these suggestions lead to more complex problems. However, the reality is that many (maybe most) people do not delete or file their email. The challenge is to explore this new space, this bit of *terra incognito*, to find new ways that work, new ways that people will adopt. That process requires some of the soft skills I mentioned earlier. A bit of anthropology, sociology, and psychology wouldn't hurt.

Continuing to explore cyberspace, archivists will quickly discover a vast area of digital preservation. Many individuals have done extensive investigation in this area, such as media longevity and format migration. Questions about how subtle changes in the way a document is rendered can affect authenticity and meaning of a record are very important and interesting. This work is invaluable. At the same time, it is often fragmented, and it is hard to see how those pieces fit together. More challenging, records professionals often find this information of little value when trying to offer recordkeepers practical advice on managing records.

Fortunately, to end on a positive note, a number of projects are trying to find ways to apply the ideas by

developing workflows. The Persistent Digital Archives and Library System (PeDALS) project is trying to automate processing electronic archival records.[11] Archivists in seven states worked together to define a common methodology to acquire, accession, describe, store, and provide access to electronic records. The methodology was expressed as business rules, discrete steps that archivists go through to curate a collection. Those business rules were then implemented in software. Writing the code took time, but it took considerably less time than manually processing the collections.

For example, the rules for accessioning records include taking an inventory to ensure that all files were received, that no extra files were received, and that the files' integrity was not compromised. The rules to describe the records include running the New Zealand Metadata Extractor to capture preservation metadata. Rules for description also include writing rules to map metadata received with the records to a standard schema.

This approach is, I believe, a paradigm shift. Archivists will no longer work directly with records. Given current resources, traditional approaches will not scale to inspect, organize, describe, and preserve a million emails. In

---

[11] Persistent Digital Archives and Library System (PeDALS). Principal support from the Library of Congress, National Digital Information Infrastructure and Preservation Program, with additional funding from the Institute of Museum and Library Services, Library Services and Technology Act. See http://www.pedalspreservation.org/ (checked 18 October 2010).

In addition to PeDALS, other projects are addressing similar concerns. Reagan Moore, Richard Marciano, and Chien-Yi Hou have been leaders in the area of distributed storage and rules-based processing through their work on iRODS and DCAPE. See "IRODS: Data Grids, Digital Libraries, Persistent Archives, and Real-time Data Systems" (DICE, 2010); https://www.irods.org/ (checked 18 October 2010) and DCAPE: Distribute Custodial Archival Preservation Environments (SALT, 2010); http://salt.unc.edu/dcape/ (checked 18 October 2010).

essence, archivists must find ways to take advantage of the fact that these are *digital* records, which lend themselves to machine processing.

These tales of cyberspace are hardly a basket of fruit and cookies from the Welcome Wagon. The profession is faced with two equally frightening realities: the vast amount that we do not yet know and the need to reconceptualize how we do our job. Dante tells us that the inscription above the gates of hell reads, "abandon hope all ye who enter here." The same might be appropriate for cyberspace.

Whenever talking about the challenges records professionals face, I fear that I will trigger paralysis, the ultimate form of denial. Instead, I would like to leave them with encouraging words, with a sense of hope. Rather than fear of the unknown, I hope through a bit of autobiography my colleagues will sense opportunity and discovery in a new and untamed land. Originally, I did not want to work with electronic records. I knew it would be a lot of hard work, although I am happy to tackle a challenge. What scared me was that I knew there was real chance of failure, and I dislike failure. Fynnette Eaton, electronic records archivist at the Smithsonian at the time, gave me the courage to dive in when she told me, "Whatever we do, we may fail. But if we do nothing, failure is guaranteed."

So, welcome to cyberspace! Dive in and give it your best! I promise you that when you do, you will find a fascinating world!

Richard Pearce-Moses has been a professional archivist for more than thirty years. He is a Fellow of the Society of American Archivists and has been a member of the Academy of Certified Archivist since its inception. Currently, he is the Director of the Master of Archival Studies program at Clayton State University in Morrow, Georgia. Previously, he served as Deputy Director for Technology and Information Resources at the Arizona

State Library and Archives. He has worked with a variety of subjects and formats, including photography, regional history, Native American art and culture, and state and local government. For the past decade, he has focused on digital archives and libraries, including finding ways to capture and preserve digital publications on the Web and new ways to automate processing electronic records.

He served as the President of the Society of American Archivists in 2005-2006. The American Library Association presented him with the Kilgour Award for Research in Library and Information Technology in 2007, and the Library of Congress named him a Digital Preservation Pioneer in 2008. Pearce-Moses is the principal author of *A Glossary of Archival and Records Terminology* (Society of American Archivists, 2005).