Why the Life-cycle of Business Records Cannot Be Managed Using Digital Technologies[1]

By Carol E.B. Choksy, Ph.D., CRM

Adjunct Professor, School of Library and Information Science, Indiana University

I am delighted to be able to address you today on a topic that is dear to me. I will address you today as a consultant who performs records management consulting for global organizations in many industry sectors including government, law firms, financial services, and manufacturing. The observations below are from my work as a consultant and informed by conversations with other consultants as well as technology companies in the information management business. None of these problems are new; they are simply exacerbated by the digital technology many believe will save us.

Technology has greatly aided the creation and distribution of documents, but technologies to manage the other phases of the life-cycle consistently do not exist. Document management technologies, scanners, and e-mail have dramatically improved our ability to create, refine, and distribute documents. Those same technologies give us the impression that we are "managing" those documents. In real businesses, these technologies have created an information chaos not previously imagined in the business world. We have been lulled into a false sense of order and authority by the very technologies we have employed.

This presentation will discuss the effects of digital technology on the creation and

[1] Presentation for *Le numérique: Impact sur le cycle de vie du document pour une analyse interdisciplinaire,* Archives National de Quebec, Montreal, Quebec, Canada, October 13, 2004.

distribution phases in order to better understand the ramifications of those technologies and subsequent real world practices on storage, destruction, and archiving. Documents are easily created. The easy distribution of those documents creates copies in many different formats and locations. That very ease of distribution frustrates any attempt to identify, maintain, protect, and preserve an authoritative copy. The lack of system for identifying an authoritative copy leaves maintenance and protection during the active use of documents in the hands and personal preferences of individual employees. The multifarious copies of every medium and format do not necessarily make their way through the destruction or archiving process. These issues create problems for archival appraisal by inundating the archives with a morass of copies in many media with every conceivable organization.

A brief example should suffice to illustrate this problem. A large, international law firm headquartered in New York discovered that some of its e-mails were being demanded for litigation on a regular basis. We were called in to assist with this problem. What we discovered was that since the inception of the electronic document repository, no documents, electronic or otherwise, had been managed. The document repository, installed in about 1995, had never been properly set up so everyone did their searches using full-text search or by client ID—there were no indices. The paper repository was incomplete because some attorneys were priding themselves on being "paperless." Some of the paper files had numerous copies of the same document because when a case or deal was over everyone sent their convenience copies to the file room. The file room did not have enough time to do anything but put all the copies in the corresponding paper file. E-mail

was managed by each employee differently and appeared to have little or no self-conscious relationship to the remainder of the repositories. This meant there were as many copies of e-mails and attachments as there were recipients. This law firm had a records management department whose portfolio extended only to paper records. None of the employees within that department had any knowledge or understanding of electronic documents or e-mail and steadfastly avoided all opportunities to learn.

Let us tease out the difficulties found here. Documents are created with great ease. Employees send instant messages and text messages (SMS) to each other and to those outside the organization without thought of how that behavior affects the management of records. Those documents do not disappear; they stay around in repositories in various organizations such as the phone company. E-mails in business are documents, but their contents resemble utterances in a conversation more than they do documents. Because of their peculiar nature, they proliferate at an alarming rate. Instead of printing a document and sending hard copies, employees now send documents as attachments to e-mails whereupon those documents are printed, read, and kept by the recipient.[i] Normal practice by most employees is to print the attachment, then move the e-mail transmittal with its attachment to a personal e-mail folder. Many persons print e-mails and keep them in a personal filing system. Others have their own e-mail filing system and keep them in their own folders rather than use public folders. The reasons given vary, but tend to center on having a reference copy ready at hand, not trusting others to keep a reference copy, or believing they will be reprimanded if they cannot respond quickly enough to the topic of the document. This means there is no “authoritative” copy, nor is there an office of re-

cord.

As you have just heard, the speed and ease with which documents are created is matched only by the speed with which they can be distributed. The two, creation and distribution, have even created new "categories" of documents such as e-mail, instant messaging, and text messaging. These documents function in the same way many old documents did, but they also serve other, older functions not previously served by documents. What used to be called "water-cooler conversation" is now written down and circulated via e-mail. Notes passed during meetings are now sent via instant messaging. Phone calls arranging meetings are now sent via e-mail as well as instant messaging. In the past, if this information was captured in some form, it usually made its way into the waste paper basket. The problem now is that this information is being captured and stored in numerous repositories, including smart phones and that may or may not be managed by the organization.

Creation and distribution being so easy, one would thing that lassoing one copy and corralling it into a specific location for life-cycle management would be a simple matter. Which copy, which version, which draft were you thinking of? Did you know that the printed Starr report submitted to Congress is not the same as the one published on the Internet?[ii] The word processing program used to create the report, WordPerfect, had an undocumented feature whereby it restored the last set of edits when it was converted to HTML. One copy was presented to the U.S. Congress, the printed version, and the other was presented to the American public, the HTML version. None of our definitions of authenticity or evidence will help us decide which of these versions is the "official copy."

Which version of the Blair government document would you have saved? Would you have saved the one with the creation and revision history that nearly brought down the government or the one that had purged that information?[iii]

What if the only copy of the document is stored at the phone company as SMS text messages are? You have no express contract with the phone company to store this information. You have no mechanism for storing the document in its appropriate format, retaining all the appropriate metadata. The phone company may even deny they keep copies of such messages.[iv]

Let us assume you could decide upon and get possession of an authoritative version of the document. The purpose of keeping an "official copy" is to assign responsibility to a particular department that will manage the remainder of the document's life-cycle. Are the members of that department sufficiently knowledgeable about the technology and the details of that document's life-cycle to truly make informed decisions? Are they aware of all copies made, their location and medium? For example, say a document is imaged for use in a business process. The organization's practice is to destroy the hard copy document and use the image as the official copy. An exception is found during the business process and a letter must be sent to the creator of the document describing the problem. The document is printed, a letter written, the letter and document copy sent to the originator, but the organization keeps a copy of the letter and printed document copy for purposes of resolving the exception. The originator replies, the reply resolves the exception, the business process continues with the image, but the reply, the letter and the document copy are kept in a folder by an employee whose performance evaluation is judged on the

basis of how well and how quickly exceptions are handled. The department managing the images is made up of members of the business process that handles the day-to-day operations using the images. Those images are destroyed after two years. The employee with the paper copies is part of another workgroup that just handles exceptions. Those exception files are kept for three years so that statistical reports can be made to discern whether the day-to-day business process is functioning adequately, or whether more of the exceptions could be handled by the day-to-day operations group.

According to your records the official copy has been destroyed during the normal course of business. Have you adequately captured the exceptions-handling records series where another copy is kept? Did you even know there was another copy? Is that other copy an official copy also?[v]

What about all those convenience copies? Have you educated your staff sufficiently so that they are destroying those on schedule? Do you have adequate motivation in your personnel management program and audit controls to determine whether you are having an impact? Your organization may not have the time, the money, or the interest *even if it is a government entity* to manage convenience copies adequately.

E-mail is particularly complicated in this regard. The way your e-mail system is set up may make it impossible to manage copies automatically. If your e-mail system is set up as IMAP (Internet Mail Application Protocol), where e-mail is delivered to a central server, there will normally be one copy of an e-mail with a number of pointers to it. If your e-mail is set up as POP (Post-office Protocol), where e-mails are delivered to the C-drive of each personal computer, there are as many copies as there are named recipients.

Attachments do not make this easier. Even with IMAP e-mail, each individual may either print or copy the attachment, making as many copies as there are recipients.

Only now are document management and content management products integrating appropriately with e-mail. The products available to manage e-mail are primarily repositories that permit only full-text search and a couple of time-based retention periods.

The problem now is, and the United States standard DOD 5015.2 is primarily to blame for this, even where integration exists, employees are not profiling and filing their e-mails. DOD 5015.2 dictates a large amount of functionality that is delightful for the archivist, but has no regard for the end-user. DOD 5015.2 dictates what shall be done, but not how it shall be done. To pass the test a product can require 25 mouse clicks by the end-user that is excluding the text fields. The end result is that end-users find ways around these products to make their lives easier. Remember that the e-mails are replacing conversations? The end-user is not being asked to profile and classify a document, but an entire conversation. And each end of the conversation is being required to do this, not just one member. Document management companies complained of this same problem in the late 1990s that end-users were doing everything possible to get around profiling entire documents. E-mail and text messaging are only exacerbating the problem.

If you knew where every copy was you might be able to create policies and procedures to destroy them. You might even be able to create motivations through performance evaluations to ensure employees perform. However, is the office of record truly knowledgeable enough to ensure a document is truly destroyed or truly preserved appropriately? The chances are that they are not. One of the problems with event-based

destruction is that those events do not always occur to the set of documents burned on to the same disk. One of the solutions is to use migration and conversion as opportunities to cull only those documents for which the trigger has not yet been pulled. In the meantime, the old ones will still exist.

As you can see, the records manager is as much concerned with destruction as she is with preservation. What often gets our organizations, both privately held and governmental, into the most trouble is not the information they have preserved, but the information they believed was deleted. Arthur Andersen is a good example.

We have examined the problem of copies in the electronic environment, but let us take a step back and look at some more technical challenges. One of the biggest challenges is that there is no product designed to manage the entire life cycle of any document. Even tactical applications that handle, for example, accounts payable can manage only part of the life cycle. Those products capture the image of the document, assist it through its business process, burn it to a disk when it is not referenced frequently, and tracked on the disk until it is time to destroy the disk. These products cannot destroy the information, nor can they destroy the original document. The entire disk must be crushed and burned—a process requiring the intervention of a human and another product or set of products.

Enterprise Content Management applications may manage information from its inception through its distribution, but what happens when the document needs to be taken off on-line or near-line storage. The product itself probably could not track the document to off-line storage. Most content and document management applications cannot track the

document to a disk or tape. Those applications certainly cannot destroy the electronic document. Many of them can track a document's retention if that retention is a simple period of time. When they delete, they do so solely in an on-line, magnetic disk environment. All they delete is the pointer, the document continues to exist until it is overwritten.

Records management applications are really just sets of business rules applied to particular content or document repositories. Those applications can certainly track the retention of a document, even if the document is off-line. Unfortunately, the only document these applications can truly destroy, not just by deleting the pointer but by over-writing with nonsense characters, is one in an on-line, magnetic disk environment.

There are many sources to these problems. One of the most pernicious is IT's lack of knowledge of documents as opposed to databases. Peruse any textbook on systems analysis and design. Read the section on business process analysis or output. The word document does not appear. I know because I teach this subject and get paid to review the textbooks. This lack of training in documents creates corporate cultures where the information technology department believes everything can be done with databases—even in industries like insurance where there are government requirements concerning documents.

There are vocabulary problems. You have heard the ones within the software industry regarding content, documents, and records. This problem exists as well among archivists and records managers. Archivists define terms such as "authentic" without regard to their U.S. records management brothers' and sisters' needs to reference the laws of discovery and document production where the term "authentic" has very different, reserved

meanings. Archivists define records as "evidence of a business transaction" without regard to their U.S. records management brothers' and sisters' needs to reference the law of evidence where the term "evidence" has very different, reserved meanings and where terms such as "business" and "transaction" are commonly used to refer to very specific activities and documents.

You may not have given much thought to the proliferation of repositories within any organization. I was recently in an international insurance agency where the director of security gave us a copy of the company's security standards. She did so using a USB drive (a "thumb drive"). I asked her if she had policies regarding such drives. She said no. Repository management is now a necessity. The first piece of advice I give to a records manager wanting to do a records inventory is to inventory all the repositories first and find out their relationships, privacy, security and role within the organization. One of the things we discover is the incommensurability of electronic repositories. Instant messages cannot be stored in the content repository and text messages are not even being maintained within the organization.

We have also found that electronic documents are privileged and hard copy documents are non-privileged unless printed by the user. This problem is worst where there is a hard-copy repository. The records manager, the only person in the entire organization who understands the life cycle, is relegated to the hard-copy repository. No one is in charge of the electronic repositories.

Technology has convinced us that we are managing documents. In fact, all technology has done is to make it even more difficult to manage our documents. For documents

Information Technology is plumbing—very good plumbing, and a copy machine, a very good copy machine. We also need to manage our understanding of what we are managing. The term "records" has been relegated either to paper or to some rarified catalog of documents meeting requirements known only to archivists. All information held within an organization gets organizations into trouble regardless of whether it is in someone's head, in a database, or a document. The information need not be part of a business process, a final draft or version or the official copy. The process of discovery in U.S. litigation has made the most inaccessible documents, e-mail: the most feared documents by compliance officers. The fear comes not only from the lack of life-cycle management technologically, but also because of the way it is used. People say just about anything in e-mail and send multiple copies throughout the organization.

Management is a practice, not a technology. What we need to manage the documents life cycle is more management, not more technology.

---

[i] Abigail J. Sellen and Richard H. R. Harper. *The Myth of the Paperless Office* (Cambridge, MA: MIT Press, 2002), 14.

[ii] See, for example, the Washington Post article by John Mintz and Nathan Abze, "Glitches Altered Starr Report Online, in Post," *Washington Post* September 16, 1998; Page A34, also at http://www.washingtonpost.com/wp-srv/politics/special/clinton/stories/document091698.htm (4 August 2004)

[iii] "Iraq-Its Infrastructure Of Concealment, Deception And Intimidation" (30 January 2003). See Glenn Rangwala's discussion of this at http://www.casi.org.uk/discuss/2003/msg00457.html (4 August 2004)

[iv] Associated Press, "Think Before You Text: Wireless Messages May Show Up in Court," CNN June 7, 2004 http://www.cnn.com/2004/TECH/ptech/06/07/text.messaging.records.ap/ (4 August 2004).

[v] You should have known that the exception-handling records series existed. Yes, the hard copy is also an official copy that is part of a larger case file within the exception-handling records series.