The Perfect and the Possible: Becoming a Digital Archivist

Richard Pearce-Moses
Presented at the Conference of Inter-Mountain Archivists
12 May 2006 : Ogden, Utah

Thank you for the invitation to speak. I'm honored. I've been looking forward to this chance to get together with you and to talk about a subject that's dear to my heart. When I was elected as President of the Society of American Archivists, I wanted to spend my term talking about the impact that technology has had – and may have – on the archival profession. Last year, when I gave my incoming speech in New Orleans, I likened that impact to a hurricane. A week later, Katrina blew through the Gulf Coast. When I speak in DC this year, I promise to somehow use the metaphor of world peace and pray for a similar coincidence.

On behalf of Southwest Archivists and American Archivists, thank you for your contribution to the Emergency Disaster Assistance Fund. I have often been overwhelmed by the generosity of my colleagues. I am convinced that the principle reason I am an archivist is because I like archivists so very much.

One thing I saw while I was visiting the coast just a few weeks after the storm was that even small efforts towards disaster preparedness could make a real difference in potential damage to records. For example, keeping records in boxes does a lot to protect records in case of a leaking roof or broken pipe. But how many people have collections sitting open on shelves? One of my hobby horses is to get archives to ensure that all their collections are in boxes from the moment they are received. I would somewhat scandalize a friend whenever I would tell folks that any box – even a cardboard box from the store with "Tide" on the side – was better than nothing. My friend argued that I shouldn't be promoting the use of boxes made of such poor materials. I stood my ground, though. Acid neutral, lignin free boxes are expensive, and many repositories can't afford them. Better to take an imperfect step towards preservation rather than do nothing because it couldn't be done "right."

In many ways, my motto has become "Don't let the perfect be the enemy of the possible." Best practices are relative, not ideal. We can always do better. We can get things in cheap boxes today, providing significant protection in case of water leaks or fires. In time, we can get better boxes. And someday, we may get state-of-the-art fire suppression as well. But we'll do the best we can now.

I think the same adage applies in the realm of electronic records and digital information.

Last year the National Archives celebrated its twentieth anniversary as an independent agency. Allen Weinstein, the Archivist of the United States, invited a number of individuals to consider how the profession has changed over that time. Bob Horton, the Minnesota State Archivist, offered his opinion that

> We have collectively experienced a technological revolution in the past decade. . . . We have not experienced the corresponding and overdue institutional and professional revolution that is the appropriate and necessary response.

That's something of a bleak observation. I suspect that one reason archivists have failed to respond is that they don't have a clear answer to a very basic question, "What should I do?"

A little over a year ago, this question surfaced when Weinstein met with the leaders of the Council of State Archivists, NAGARA, and SAA. Weinstein proposed a colloquium to bring together archivists with experience working with electronic records; this group would try to develop some answers to that basic question. I'm happy to tell you that the colloquium, sponsored by the National Archives and Records Administration (NARA), the Society of American Archivists (SAA), and the Arizona State Library and Archives will take place at the end of the month. Proceedings from the conference will be published and should be on SAA's website this Fall.

At the same time, the Council of the Society of American Archivists has been trying to articulate critical, strategic issues the profession is facing. What's on the horizon that could have a significant impact on who we are and what we do as archivists? The impact of technology was at the top of the list. Council's conversation about the role of technology included a wide range of viewpoints.

+ Some felt that technology threatens the very future of the profession. They could envision a digital future where archivists were relegated to the realm of paper, and information technologists would care for the digital record. This point of view may seem radical, but there is some evidence to support it. And while it's by no means a foregone conclusion, we must address the possibility head on. It's easy to assume that archivists have a skills set that ensures our role in the information ecosystem, but we do so at our peril. For example, one archivist that I have the greatest respect for commented to me once that IT people don't appraise records; that's what archivists do. I challenged her with my own observation that a technologist's decision to delete all email after ninety days is an appraisal decision; we may not think it's a good decision, but in some organizations that's the reality.

+ Others felt that the profession has made great strides and that we should be more optimistic. I agree that archivists have made progress on many fronts. One of the first things archivists had to do was learn about the very nature of the beast. I think most people in literate cultures forget that pen and paper are, in fact, also an information technology. It's so familiar that it's essentially transparent. We are surrounded by this technology, and we learned about it by observation rather than critical analysis. As such, our knowledge of this analog technology is, to a large extent, intuitive rather than empirical. We must become as familiar with digital formats as we are with paper formats.

  Some leading lights in the field have done significant research into the nature of electronic records. A lot has been published on the authenticity, reliability of electronic

recordkeeping systems, and preservation of electronic records. Interestingly, in the process, we have seen two major works that help us better understand paper technologies. I commend to your reading pleasure and professional development Sellen and Harper's *The Myth of the Paperless Office* and JoAnne Yates' *Control Through Communication*.

No doubt, the reality is somewhere between these two points of view. Neither should we despair of dire pessimism nor rejoice in Pollyannaish optimism. We have made progress, but the fact that we've come a long way doesn't discount the fact that we still have a lot further to go. I also believe that, given the pace of change, we must redouble our efforts, or we may be left behind.

Which brings us back to the question, "What should I do?" Last week I was talking with Peter Gottlieb, the Wisconsin State Archivist, about this very question. He observed that the only way to really learn how to work with these materials is to just dive in. I think his comment hits the nail on the head.

Simply put, do *something*. Or, as the adage puts it, a journey of a thousand miles begins with a single step.

If you know nothing about electronic records, I encourage you to start learning. My column in the forthcoming issue of *Archival Outlook* lists some basic readings. In addition to the works I mentioned above, I'd highlight David Levy's *Scrolling Forward* and Bryan Bergeron's *Dark Ages II: When the Digital Data Die*.

I offer a few words of caution here as you begin.

+ Much of the archival literature on electronic records is very theoretical. I believe it's essential to have a strong grasp of that theory, as that knowledge guides our actions. At the same time, it's critical to use that knowledge to get the work done. Practice requires applied knowledge, and the best place to get that applied knowledge is the school of hard knocks.

+ Don't look for a magic bullet. As far as I can tell, there's no single solution to the problem of electronic records. There are many kinds of electronic records, and no one approach will work for them all. Some approaches fail to scale; what works well for a large series may not be practical for a smaller series. Some approaches don't translate from one administrative context to another; an approach that assumes a organization with a centralized IT shop may fail in another agency with decentralized IT.

+ Don't think too far into the future. Technology is changing too fast to plan for more than three to five years out. What looks like an insurmountable problem now may be moot, and continued experience with electronic records gives us new insights and new solutions.

In some ways, the decision to *do something* changes the question from "What do I do?" to "What do I want to do?" I would like to think that archivists' appraisal skills should enable them to answer this question fairly easily. Which records are at great risk and of significant value?"

+ Are the records truly at risk? Some of the most important records are still found in traditional paper formats, so the need to preserve the electronic version is reduced. We may want to find some way to capture electronic versions, but they might not be my first choice if I knew those records existed in paper.

+ Are the records of equal value, or critical value? The records of the Governor are almost certainly more important than those of the Barbers Commission. Within an office, policy decisions and records documenting rights and entitlements are almost certainly going to be more important than email correspondence. (This is not to say that the records of the Barbers Commission are not important or that there aren't email messages that merit archival preservation. It's a matter of relative importance and priority.)

+ Do the records justify the resources necessary to curate and preserve them? An archives might pass on a collection of nitrate negatives because it does not have the facilities to duplicate them on safety film or to put them in cold storage. Similarly, an archives might start with a collection of e-records that it has the resources to manage. It takes a lot less work to preserve a simple collection of static documents than a complex geographical information system.

So that's the ideal based on the principles of appraisal. In reality, you may not start with records of great value or at significant risk. You may very well find yourself starting with a somewhat arbitrary collection of records because of some bizarre circumstances. If a CD with an important database gets thrown over the transom, start there. Don't think this to the point of paralysis; the goal is to get started, not to find the best records with which to begin. Take advantage of strategic opportunities. If nothing else, work with the electronic records on your own PC.

Now that you've decided to jump in and you've selected a collection to work with, you're still faced with the question. "What do I do?" At one level, you know the answer to this question. You need to do the same things with this collection of digital records that you need to do with any collection of records. These activities are the core functions of our profession. The question becomes, "How do I do this?"

Appraisal and Acquisition. In the digital era, we won't use boxes to acquire records. We need some other sort of container, such as disk or tape. But in the digital era, we have another option: networked file transfer. We need he ability to use a variety of techniques to select records from a creator's system and the ability to transfer them to the archives' system,

including the use of physical media and the network. This might include knowing how to burn a CD or to ftp files.

Accessioning. In the digital era, we still need to document the acquisition of these records. We may be able to continue using existing techniques here, as this is a record and recordkeeping system we're created ourselves. At the same time, we should be taking advantage of the power of technology, using electronic recordkeeping systems for our own purposes. We need the ability to develop simple databases to improve work flow and recordkeeping.

Processing (*Arrangement and Description*). When we acquire paper records, we check for pests and mold, organize them, we stabilize and rehouse them, and we create finding aids. In the digital era, we will run virus scans. We may need to transform them from a native format to one that we can more easily preserve, such as XML or PDF. And, we still must describe the records so that our patrons can find them. Here, theoretical knowledge is particularly important so that we do the right thing; for example, we have to know the long-term implications of our actions, such as choosing PDF, XML, or something else. We need to know file formats, as well as processes to sort and transform data. As I just mentioned, we need to know how to exploit technology to help us do this better: in an age when patrons are accustomed to full-text searches on the web, they will not likely be satisfied with a hierarchical description of e-folders. We need to know how to use technology to help us do our jobs better.

Preservation. When we've received fragile or unstable materials, we often undertake a project to preserve those records. We microfilm acidic newspaper and books, and we print or copy nitrate negatives. When completed, these reformatted records would remain stable for years if properly stored. In the digital environment, preservation will no longer be a project, but a process. For the foreseeable future, we will not convert records to a permanent format. Any format we choose will eventually – and seemingly sooner rather than later – become obsolete. At that point we will have to migrate the records again. From CD to DVD to the next new media. From WordStar to WordPerfect to Word to the next new document format. We may need to learn batch processing for massive conversion projects. As likely, most of us will need to learn project management skills to oversee this type of work being done by contractors.

We need to hone a variety of skills. In addition to technical skills, the ability to work with people is critical. Knowing how to talk to a technologist – how to build an alliance and sympathy with someone who is really busy – may be the key to exporting records from a critical system.

But technical skills are the heart of the matter. We need to know about technology to be able to talk intelligently with systems administrators. We need to know the limits of technology to set reasonable expectations and to manage technical projects.

I think a lot of archivists are afraid that they need to learn programming. I have to tell you, I am very grateful for knowing a little about programming, database design, and systems

administration. But I don't want to suggest that I'm anything but an amateur. I couldn't ever get a job as an IT professional. But I know enough of the basics to have prototyped tools to help manage digital information. That's been invaluable to help me understand exactly what needs to be done. As a result, I've had a real leg up when talking to technologists, and I have a sense as to whether their work is on track. I often encourage people to take a basic programming course at a local community college or to take an old PC and install the Linux operating system and getting the Apache web server running.

While archivists may not need to know programming, they must know more than how to use basic desktop applications. Archivists need to be able to do more than create and edit documents or to build spreadsheets and databases. Those are the activities of records creators. Archivists need the skills to manage those records. At a minimum, that means a knowledge of moving, renaming, sorting, selecting, and describing files. On a more advanced level, we will need to perform these operations on thousands of files. At a minimum, we will need the ability to write simple scripts, and in some cases this may blur into the realm of programming.

When I encourage people to take a programming course or play with Linux and Apache, I don't expect them to become master programmers or systems administrators. I hope for two things. First, I hope they will begin to become comfortable in what for many is an alien world. Second, I want them to become more comfortable with risk taking and failure. It took me several attempts to get Linux running. Experimenting with technology helps one discover the real limits of loss. Because I had the original installation disks and I was working on a test system, I could try anything. I could crash and burn the system without fear, and I began to learn what would (and wouldn't) do that. I learned other things, too; after innumerable attempts to load the software, I discovered the problem wasn't me or the software, it was a defective hard drive. How would I have ever learned that lesson without experimenting?

Some people envision archivists who specialize in digital records, similar to those who specialize in photographs or sound recordings. I don't. I think that essentially all archivists will need those skills because all archivists will – with rare exceptions – work with digital materials. Some people think I'm a bit extreme here. Some archivists – usually managers – tell me they don't need this technical knowledge. As often as not they tell me, "I'll hire those skills." They may be right. But I believe that we need archivists with technical skills to do archival work, not technologists with some archival skills. Archivists bring core values and principles to their jobs; a hot-shot programmer may not be aware of how their decisions threaten the reliability, authenticity, or long-term preservation of the records.

It's not enough to focus on the new skills for the digital era. We also need a new attitude. WWW really stands for the 'wild, wild web.' Electronic records are a new frontier, and it's not for the faint of heart. We need archivists who are early adopters; people who are excited, rather than intimidated, by new technology and innovations. We need risk takers; people who are willing to try something new, and when confronted with failure keep trying it over and over

until they master it. We need people who are problem solvers and committed to finding creative solutions.

Most important, we must not let the perfect be the enemy of the possible. We need people who will focus on what we can do. Let us celebrate the reality of what we *can* accomplish, rather than bemoan the dream we did not fully realize.

If I can emphasize one point today, it's this. This is new territory. Whatever we do, we may fail. But if we do nothing, failure is assured. I don't say this with despair. To the contrary, I tell you this as comfort to encourage you to begin. Our fear of failure can paralyze us. Early adopters, risk takers, and creative people are comfortable with failure. It's not that they like to fail; they often take failure very hard. But they know how to pick themselves up and start again.

I'm encouraging you to take a sink-or-swim approach to electronic records. Dive in. Start thrashing around, and keep doing the things that help you stay afloat and move forward in the water. But don't go swimming alone. Use the buddy system. Use organizations like CIMA and SAA to build a network of friends who can give you tips – and consolation when needed. When I first started working with electronic records, I was overwhelmed. I wasn't sure what to do, and some days I feel clueless. I was afraid of failure, and frankly I still am. That's when I touch base with the good friends I've made in the field. And every once in a while, when I feel overwhelmed by all that lies ahead, I look around at how far I've come: what I've learned and the progress we've made in Arizona.

Today, I've emphasized new skills and new attitudes. However, I firmly believe that our knowledge of traditional archives remains invaluable. I am convinced that what we do remains the same; it's only how we do it that will change. For every twist and turn I've seen in the world of electronic records, I've seen an analogous quirk in paper that offers insight on how we should approach the solution. The future needs archivists. We have to be ready to step up to the plate.

Take courage, take heart, and welcome to the digital era.