

Dan Santamaria

Seeley G. Mudd Manuscript Library, Princeton University

Remarks for panel on "Reconciling Modern Archival Practices and Ethics with Large Scale Digitization"

The Legal and Ethical Implications of Large-Scale Digitization of Manuscript Collections Symposium

Southern Historical Collections, University of North Carolina, Chapel Hill

12 February 2009

My fellow panelists have addressed specific strategies for approaching large scale digitization. My role is to focus on some of the broader ethical implications of this type of work (or perhaps the implications of failing to embrace this type of work.)

I'd like to start with an anecdote, in hopes of illustrating a user prospective. – A few months ago my first child was born. One evening, a few weeks after his birth, my wife went to check the mail and was shocked to find a notice about an abnormal result on a newborn screening test. I won't go into details, except to say that everything turned out fine, but this event led to several days of phone calls, difficulty tracking down information, and of course a whole lot of questions– why hadn't our doctor, or anyone, received the actual test results, why in this day and age, did the notice come in the mail several weeks later, why was there almost no information about what steps we should take or who we could contact. At one point, after a day and a half, my wife – who is also an archivist - called the new jersey health department to track down the results, and was told it could take several days to obtain the records we needed. When she informed the employee that multiple people had been trying to get the information for several days – she was told "you don't understand – we have thousands of documents here so it takes a long time to find things."

I obviously complained and related this story to various people at work, but after a few retellings I realized how similar the employee sounded to our own staff – yes, even me at times – when talking to our patrons.

I've noted, as I'm sure many of you have, increasing incredulousness as I attempt to explain our access policies to patrons – why is very little available online or in any electronic form? why do they need to travel to the library to conduct research? why can't we just email them the photocopies they requested 4 weeks ago instead of making them pay for shipping?

Our traditional responses to these questions – the cost of digitization, staffing resources, digital preservation issues – are increasingly falling on deaf ears. Our users expect to find information quickly, at whatever time they happen to need it.

I'm certainly not the first or only person to notice this trend.

So what does this all have to do with ethics?

The SAA code of ethics states:

"Archivists strive to promote open and equitable access to their services and the records in their care without discrimination or preferential treatment..."

One of the key concepts for me is equitable – or equal- access. If we continue our usual procedures we are limiting access to those who can afford to travel to our repositories and in most cases those who can afford to come during working hours. This isn't really equal access. It's access to an "elite few" (credit Max Evans for "elite few" phrase from PACSL conference)

Gerald Ham's writing about appraisal, mentioned frequently by Mark Greene in recent years, particularly in MPLP and his presidential addresses, certainly seems to apply here:

Ham wrote about archivists' reluctance to "make hard appraisal decisions and to instead request ever more storage space and said 'Society must regard such broadness of spirit as profligacy, if not outright idiocy.'"

Doesn't this also apply, today, to storing and preserving records that much of the population views as inaccessible? How can we continue to advocate for additional resources from our parent institutions and generally justify our importance if we meet almost no one's expectations? The current state of the economy only serves to emphasize this point.

Privacy and confidentiality are important and legal restrictions and restrictions as a matter of established policy are a necessary part of our work, but we cannot let them trump all other considerations. As Mark Greene has written, "We must get beyond our absurd over-cautiousness that unprocessed collections might harbor embarrassing material not accounted for in deeds of gift." If good faith efforts have been made to contact donors collections with no deeds of gift should also be candidates for digitization; otherwise they should be candidates for reappraisal and deaccessioning.

We have, I would argue, examples of what happens when we go down this road – of essentially hunting for problems – with government declassification. I'll use an example from my own repository – the Allen Dulles papers, which were seized and reviewed by the CIA before coming to Princeton. Just last year redacted documents were finally delivered to us, which revealed that important national

security documents such as time magazine articles and blank calendar pages had been restricted for decades. Hunting for ambiguously defined confidential or embarrassing material has also been a contributing factor in our collective slow processing rates and large backlogs.

The title of this panel is reconciling digitization with current archival practice. Our hesitance to digitize strikes me as very reminiscent of our reluctance to let go of some of our traditional processing practices. I teach a workshop on pragmatic processing strategies, and at each workshop I still hear a number of objections to "MPLP" informed processing. Whether specifically focused on preservation or description or privacy and confidentiality, the common thread with all of them is a sense of losing control. I would like to think that archivists today, as emerging technologies are making all sorts of repurposing of resources possible, would think that it is exactly our job to provide access to collections and let patrons use and repurpose materials in whatever legal ways they want, but it seems to me that we, as a profession, are very reluctant to do so.

At SAA this year I heard a number of comments about loss of control in sessions about digitization. My favorite response came from Beth Yakel who noted that "we really don't have very good control of our collections, anyway..." – since most of us have long since given up the idea of cataloging individual documents and many of still labor with large backlogs. Is digitization really all that different from, say, letting a researcher take a digital photograph from an unprocessed collection in a reading room?

[I also recommend listening to Clay Shirky's talk at the recent Smithsonian 2.0 conference – using the example of content from a Smithsonian site devoted to African American history which was used to in the construction of a racist website – he argues that society has reached the point when most people differentiate between content providers and those that repurpose content. No one, he says, is holding the Smithsonian responsible for the racist website. According to Shirky the value of the Smithsonian, and of all of our repositories, lies in our ability to provide content and tools that allow people to connect with each other, not to control and mitigate access. It's a compelling vision.]

We have a fundamental ethical responsibility to provide access to our holdings. We can not fail to do so because of our own anxiety and fear - of mistakes that might be made, of what people will think, of trying to account for every possibility, of giving up control over our records and documents.

Yes, there are complications and many challenges involved with large scale digitization but we need to, and can, as I think the panel has shown, find ways to address the issues as a matter of policy. Just as with processing, in the end we

need to see access – and now digitization - as the default. Exceptional situations will surely arise, but these must come with strong arguments for any additional investment of time and effort in identifying individual problematic documents or for any additional access restrictions. [At Princeton we have developed tiered processing levels which prescribe various level of treatment for the arrangement and description of paper records; it seems to me that this approach would be successful with large-scale digitization as well...)

Our access policies have evolved over the years. I work at an institution that, until the 1990s, relied on a curatorial model – even in the University Archives – in which the curator alone would determine whether a researcher was a qualified scholar – and even then access was often dependent on whether the curator liked and was knowledgeable about a topic. I hope most of us would see the ethical failings in this approach, but I fear that increasingly our users see us as this same type of gatekeeper. We often ask ourselves if we will reach a time when if information is not online, it is does not exist. Most of our users think that time is already here.
