

Panel: Reconciling Modern Archival Practices and Ethics with Large-Scale Digitization

Extending the Reach of Southern Sources: Proceeding to Large-Scale Digitization of Manuscript Collections
Southern Historical Collection || University of North Carolina at Chapel Hill || 12 February 2009

Landis remarks (Aikens, Evans, Landis, Santamaria, Hyry -- Proffitt)

I want to start with a couple of quick observations that I've made as, over the past several weeks, I've engaged in discussions with a variety of colleagues in preparation for this panel.

- First, when archivists talk about privacy and confidentiality issues, I think it collectively brings out our most conservative streak. I typically see what I'd characterize as downward-spiraling scenarios of privacy-violation horror into which we *talk ourselves*. I would argue that this worst case-driven focus on the outliers is similar to the unproductive navel-gazing that impeded this profession's embrace of descriptive standards and shared data structures for the better part of the latter 20th century? Surely, in our vast and varied archival holdings, there are places to start that would allow us to build professional consensus and celebrate broad, unfettered access to the treasures buried therein?
- Second, we keep talking about digitizing collections as though there isn't an enormous train of born-digital content barreling down on us. I take it as a core assumption that we shouldn't be making decisions about access to *digitized* content as though somehow the issues are different from access to *natively digital* content. Worrying too exclusively about access issues for digitized content, I think, puts us in a bit of an ostrich-like position, heads buried deep in the sand, refusing to look at the ethical issue we may be ignoring by failing to capture records that individuals and organizations are creating digitally *as we speak* so that they'll be available in archives a century from now.

With that as a preface, I'll turn to the topic of donor conversations and agreements, for which I am responsible on this panel.

Regardless of what we do with the donor agreements we've inherited from our predecessors (whether last year or decades ago), the harsh and simple fact is that every year in which we don't explicitly deal with the topic of this symposium in the discussions we have and the agreements we sign with our donors, is another year of pain and ambiguity we bequeath to our successors. This might be characterized in some circles as the height of professional irresponsibility.

I would argue that thinking about current donors, and exploring what we ought to be doing with them, might help us develop a framework in which we can make better-informed decisions about digitizing in a less boutique way the collections we've had for decades, if not longer. In talking to donors now, I think we should, at minimum, be doing two things:

- First, telling them explicitly that our mission involves making the records they donate in all formats as widely available as possible, including providing access to the records themselves through ubiquitous web-based search engines, especially when they come to us already in digital formats.
- Second, clearly explaining the issue of third-party privacy in a non-threatening way, and telling them explicitly that it is their responsibility as the donor to identify materials or issues in the materials they are donating that raise concerns for them.

In addition to dealing with the individuals who give collections to our repositories, we need to be engaged as a profession in developing best practices for these discussions in a variety of institutional contexts. We also need good examples of boilerplate language to include in our deeds of gift whereby donors acknowledge that they understand issues regarding third-party privacy and their obligation to identify materials of concern to them. When a donor is not the *creator* of the records, and therefore potentially less familiar with their content, we need scenarios that take a more proactive approach to eliciting information about third-party privacy and

confidentiality concerns, and reflect this appropriately in the agreements we make with them.

Finally, I think we need to use for the ethical and legal issues in privacy and confidentiality the approaches our profession has used in educating archivists about making intellectual property risk-management assessments. A good collective understanding of viable remedies, such as access-restriction windows that might mitigate different categories of third-party privacy concerns, would go a long way towards helping us in our day-to-day work. Identifying those in our profession who can become the Peter Hirtle or Bill Maher of third-party privacy in archival continuing education and professional literature will be immensely useful to all of us as we grapple with these issues.

As archivists, we define creators as those individuals or organizations that organically accumulate, maintain, and use records in the course of doing their work or living their lives. Identification of potential third-party privacy and confidentiality issues in archival collections is arguably best vested in the creator, and the further one gets from the creator the less likely that a credible job can be done of knowledgeably identifying these issues. I think it follows that as archivists, working within the current parameters of our professional practice, we simply don't have the tools to do this work well, and no amount of funding will put us in a position to do so. With that as a governing assumption, how can we approach digitization of existing collections in a repository like the Southern Historical Collection, or in my case the manuscripts part of Manuscripts and Archives at Yale? Without taking on the insurmountable and access-dampening burden of reading every document we seek to digitize, I think we have to come up with reasonably comfortable steps to help us in doing the best possible risk analysis that we can. I'll close my comments with what feels to me like a comfortable set of such steps.

Step 1: We can discuss our digitization plan with the creator or donor, living family members, or individuals who work for an organizational creator, if these can be identified. If they express concerns, we should probably put off digitization of the parts of the collection for which their concerns are relevant.

Step 2: If the collection we're considering digitizing has been accessible to researchers who physically visit our reading rooms, we can talk to anyone who may have used it intensively. Researchers' input on potential third-party privacy issues may be helpful to us in deciding what we can feel comfortable digitizing.

Step 3: We can use existing frameworks to help us identify potential privacy issues in the records. For instance, if the creator was involved in health care issues, HIPAA may be a framework for determining how long we should be waiting to digitize. If there are materials related to personnel issues, perhaps there are state or institutional regulations or guidelines that are useful? In California, for example, a 75 year restriction would be a useful guideline for thinking about digitizing anything related to personnel issues.

It seems absolutely essential to me that we work within our professional organizations to determine what all the relevant steps might be. Once I'd taken steps like these, and assuming that researchers are already able to use the materials in my reading room, I would feel pretty confident that I had good grounds on which to champion broad access and digitize the materials in question.