

Barbara Aikens
Archives of American Art, Smithsonian Institution

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

Reconciling Traditional Archival Processing and MLP Strategies with Large Scale Digitization of Manuscript Collections

Background on Digitization at AAA

After Laura’s overview of the SHC this morning, I recognized many similarities between AAA and SHC: AAA has circa 4,500 manuscript collections, totaling over 15,000 linear feet. We also have close to 3,000 oral histories.

At AAA, we are scanning on two levels – large scale digitization of entire collections and selected individually cataloged highlights from collections in full color high resolution. For the large scale digitization of entire collections, we scan entire manuscript collections in gray scale mode at 300 dpi and present the images online as a collection of mostly thumbnail images which can be enlarged for full viewing – with watermarks. The EAD finding aid data is repurposed to serve as 1) the only metadata for the digitized collection, and 2) the online navigational and access tool.

Since 2005, we have fully digitized 65 collections, 7 of which were from existing microfilm. This totals 600,000 digital images, representing 450 linear feet of manuscript collections, about 300 feet of which were digitized from the originals. The project relies on an archival processing staff of 3 (plus myself as supervisor), 2 scanning technicians, a computer programmer, two additional specialists, and the project manager (who also wears several other hats at AAA).

Our website includes a 3 page detailed copyright and use statement. We operate under the assumption that our content constitutes fair use and the majority of it is not presented online in high resolution, and, thus can not be published by outside users. If “issues” arise, and a few have, we either remove the image, or only make the thumbnail version available.

I am providing this background to show that we are currently implementing a program for the large-scale digitization of our collections, based on our primary goal to increase researcher access to our holdings. It appears that our current project has the same goals as SHC.

Ethical Selections at the Processing Level of the Archival Workflow

Most often, the selection or prioritization of a collection for full digitization is determined by the overall research significance of the collection, much the same way that a collection is

prioritized for full processing. These overall broad appraisal and prioritization decisions are made by the curators, the director, or the head of reference or processing. Broad and general ethical, sensitivity, and legal issues are addressed at the collection-level as well. I am not going to address these broader issues here – others on the panel will be doing that today.

What I would like to focus on is the role of the processing archivist in the final selection of materials to be scanned or not scanned within a collection after the collection has already been prioritized for full digitization.

Traditionally, processing archivists have made on-going appraisal decisions while they process individual collections. Even though a collection has been prioritized for full digitization, not every single item must be scanned – and this includes materials within the collection that might be sensitive, have questionable copyright issues, or are simply inappropriate in an online environment – such as sexually explicit materials.

Processing archivists have always weeded duplicate and non-archival materials from collections, and identified and isolated sensitive materials as they process. This role or function does not change in an online environment and these intuitive appraisal decisions made by the processing archivists can be applied to large scale, collection-level digitization projects.

In our workflow, the processing archivist is the one who prepares the collection for scanning and the metadata for the digitized collection. As they process, they still think about 1) who will use this collection; 2) what will users want most from this collection; 3) what is unique, rare, or aesthetically pleasing (or not) about this collection; and 4) how can I best arrange and describe this collection to increase access? Perhaps with some simple written guidelines to assist them in their work. I have written a set of very simple procedural guidelines for our processing archivists to follow when preparing entire collections for scanning. To quote, in part, from these existing guidelines.

- If a collection is scheduled for scanning, most of the material within the collection should be scanned. The decision not to scan materials should NOT be made on an item-by-item basis, but rather on the folder and/or series/subseries level. However, not everything can or should be scanned in a collection.

The processing archivist should also look at the collection or portions of the collection for potential copyright and trademark issues, as well as inappropriate or sensitive materials. It is not the processing archivist's job to determine who owns copyright, but to be aware of the issue and discuss it with the Project Director or Supervisor if anything might be questionable.

In addition, do not have inappropriate materials scanned, such as pornography, nude pictures, or lascivious materials that are clearly not source materials for the artist or that lend nothing to the understanding of the life of the creator, such as

private letters of sexual fantasies or similar journals. Even some avant-garde artistic materials might be reviewed for appropriateness.

Also, be aware of privacy issues, such as social security numbers, mental or other health records, tax records, banking records, family histories, etc.

Within these same written guidelines is an actual list of specific documents and types of materials for the archivists to consider.

The processing archivist makes “not scanned” or “partially scanned” notations at the folder entry for materials not scanned, and a more general note in the Alternative Formats note in the EAD finding aid. Thus, the user can readily see the existence and availability of any materials not scanned.

However, I should also point out that the Archives does have many nude photographs on its website, such as artists’ models, and other non-sexually explicit nude photographs. Because we collect the papers of artists and others that may be considered on the fringe of traditional society, we do have many non-traditional materials within our collections that some may consider offensive, or that may not be readily understood by general audiences. While there is no easy answer to this dilemma, the processing archivist is often the best qualified person to decide what is the most important material within the collection for scanning.

Does this Approach Contradict MLP/ Minimal Level Processing?

I like to refer to our approach to large scale digitization as MLD – Minimal Level Digitization. However, MLD requires a somewhat more detailed level of processing. Does this contradict MLP altogether? No, the theory of MLP proposes that some significant collections will always be prioritized for full-level processing. These are probably the same collections that would be prioritized for full digitization. We will never be able to digitize ALL of our collections, just as we will probably never be able to fully process ALL of our backlog ??

Full digitization of collections may indeed require full processing – I am NOT talking about item-level processing, cataloging, or metadata though. But I am talking about traditional processing at the folder and series/subseries level.

By formally incorporating the review for items of a sensitive nature – or with copyright, legal, or privacy issues – into the work of the processing archivist, who is most likely to be in tune to these issues anyway, we have an existing well-qualified resource for indentifying sensitive materials prior to scanning.

I urge us to take advantage of this resource and to eliminate yet another barrier to implementing large scale digitization projects and programs in our repositories.