

Provocative Ideas from SAA, DLF, and SGA

Dan Santamaria
December 14, 2007

Digitizing Archival Collections

**Shifting Gears:
Gearing Up to Get Into the Flow**

OCLC/RLG FORUM

Digitization matters: Breaking through the barriers—scaling up digitization of special collections

<http://www.oclc.org/programs/events/2007-08-29.htm>

(audio available)

1. Access vs. preservation—Access wins!

- Digital initiatives have stressed the importance of preservation.
- Mass digitization of books - special collections marginalized.
- Mantra of “we’ll only get one chance to do it, so it’s got to be done right” but experience has shown that that is not in fact the case.

2. Selection has already been done

- We've already done the selection. Stop making decisions. If it's not valuable, why do we have it?
- If it's useful, researchers of all stripes may want it -- but only if they know it's there.

2. Selection has already been done

- There are three reasonable means to determine what to digitize:
 - A. Scan as materials are accessioned.
 - ◆ i. For collections known to be of great interest and that lend themselves to digitization, think about scanning the entire collection rather than making decisions about which bits.
 - ◆ ii. Otherwise think about scanning the first chunk, or a box of correspondence, or samples from several parts of the collection.

2. Selection has already been done

- There are three reasonable means to determine what to digitize:
 - B. Scan on demand.
 - ◆ As materials (whether a single item or a boxful) are requested for reading room use, circulation, reproduction, or interlibrary loan, digitize them and make the digital versions available to everyone, by dropping them into the collection- or series-level descriptions.

2. Selection has already been done

- There are three reasonable means to determine what to digitize:

C. For collections receiving minimal description or that are in the backlog, scan signposts and then devote more effort as use and interest warrant.

- ◆ “Expose tips of the icebergs, by means of multilevel descriptions and minimal processing”
- ◆ Basically Greene/Meissner for digitization.

3. Do it ONCE (then iterate)

- Do the scanning as an integral part of the initial accessioning and processing, rather than taking extra time to make another pass.
- High-level description with some representative scans may be enough.
- Once you've got all your collections represented on the web, you now have a basis upon which to determine where to apply further effort.

4. Programs not projects

- Early digitization projects have focused on cherry-picking for topical projects.
- Result in vast collections represented by a relatively small number of gorgeous images, lovingly selected, described, and presented in deep web portals.
- Digitizing our rare and unique materials must be integrated into the operating budget.
- All other important processes have budget, staffing and infrastructure, why should digitization be different?

5. Describing special collections: Take a page from archivists

- Stop obsessing about items.
- Everything that is digitized does not need to be painstakingly described.
- Archival control distinguishes organic collections from description of distinct books and museum objects.
- Let's embrace that collection management strategy as well as the standards and practices of managing collections and hierarchies, not necessarily items.

5. Describing special collections: Take a page from archivists

- Stop obsessing about items.
- We should consider opening our finding aids and other descriptions for user contributed amendments and commentary.
- We keep treating our bibliographic records like catalog cards and our finding aids like type-written documents.

6. Quality vs. quantity—Quantity wins!

- Vast quantities of digitized primary materials will trump a few superbly crafted special collections.
- Minimal description will not restrict use as much as limiting access to those who can show up in person.
- We must stop our slavish devotion to detail; the perfect has become the enemy of the possible.

6. Quality vs. quantity—Quantity wins!

- Once we know which materials are in demand, we can devote our efforts to more exhaustive scanning, higher quality scans, enhanced metadata, and text conversion -- on the content that will most benefits our users.

7. Discovery happens elsewhere

- We need to stop thinking of our lovingly crafted sites, designed specifically for a particular collection, as the only way people will discover our content.
- Our emphasis should go to making our content harvestable, collectible, and indexable by others who are more successful at reaching broad audiences.

8. Brother can you spare a dime?

- Getting grant funding in the past seemed to require identifying a “sexy” collection and maximizing its specialness in the ways we described, digitized, and presented it.
- Money went toward creating barely visited web sites. Institutions created preservation-quality images that they, in fact, had no way to sustain in the long run.

8. Brother can you spare a dime?

- Grant-giving agencies should be encouraged to support projects that put permanent processes in place for ongoing operations.
- They might follow NHPRC's lead in sponsoring initiatives that help to establish scalable processes that can be used again and again.
- We all need to stop thinking in terms of individual projects and instead think in terms of ongoing mission-essential operations.

Relevant Work in Other Repositories

- University of Wisconsin
 - ◆ Testing “More Bytes, Less Bite: Cutting Corners in Digitization.”
 - ◆ 79% reduction in the costs of putting manuscript materials online.
 - ◆ http://www.midwestarchives.org/2006_Fall/presentations/Ranger%20mahapresentationranger.doc

Relevant Work in Other Repositories

- Social tagging and user-supplied metadata
 - ◆ Michigan Polar Bear Project:
<http://polarbears.si.umich.edu/>

Relevant Work in Other Repositories

- Creating shareable metadata and repurposing data
 - ◆ Following standards
 - DACS
 - My presentation at SGA:
<http://www.princeton.edu/~dsantam/presentations/SGA11.2007.pdf>
 - EAD
 - ◆ Occasionally, not following exacting standards

Relevant Work in Other Repositories

- Creating shareable metadata and repurposing data
 - ◆ Following standards
 - EAC
 - See NCBIIO site:
<http://www.ils.unc.edu/digitalnc/ncbhio/index.htm>

Relevant Work in Other Repositories

- Implementing new tools and procedures to increase efficiency
 - ◆ “More Product, Less Process”
 - ◆ Prelinger Library (Film and Video)
 - ◆ Institutional Repositories
 - ◆ Digitool for electronic records at the Lib of Virginia
 - ◆ Digital cameras in the reading room

Relevant Work in Other Repositories

- Implementing new tools and procedures to increase efficiency
 - ◆ Archivists' Toolkit
 - My presentation at SAA:
<http://www.princeton.edu/~dsantam/presentations/SAA.8.2007.pdf>
 - Other SAA presentations:
<http://www.archiviststoolkit.org/overview/presentations/ATatSAA-2007.ppt>
 - ◆ Archivists' Toolkit v. Archon
 - ◆ Other institutions developing their own tools:
University of Tennessee MODS editor

Relevant Work in Other Repositories

- Developing programs and infrastructure
 - ◆ Fedora and DSpace have evolved into new non-profit organizations.
 - ◆ Focus for these new organizations is sustainability and governance.
 - ◆ Learn from start-up companies by embracing failures, and if one initiative does not work, shut it down and redistribute the resources.

What can we do here?

- We are in good shape and ahead of the game, in some ways (all collections represented at collection level)
- Now need to focus on the next steps...keep processing, but also figure out better ways to let people interact with and use our stuff.
- Whether it's more "interactive" finding aids, or actual digitization, etc.
- Fix problems and build a sustainable program and infrastructure.

What can we do here?

- Digitization
 - ◆ Find ways to do it ourselves
 - Patron demand
 - Digital cameras

What can we do here?

Description

- ◆ Find ways to reuse/convert the data we already have
 - Online DBs
 - Link digital objects to finding aids
 - Make use of authority data in finding aids and MASC - EAC
 - Use (and improve collection of) reference/use data to determine priorities
 - Monitor developments in the community and implement relevant tools: Archivists' Toolkit
 - Consider grants for developing programs/collaborations projects (EAC)
 - Keep processing

Thanks.

Questions comments:
Dsantam@princeton.edu