# Scan and Deliver:

# Managing User-initiated Digitization in Special Collections and Archives

## Jennifer Schaffner
OCLC Research

## Francine Snyder
Solomon R. Guggenheim Museum

## Shannon Supple
University of California at Berkeley

OCLC™

Scan and Deliver: Managing User-initiated Digitization in Special Collections and Archives
Jennifer Schaffner, Francine Snyder, and Shannon Supple

Please direct correspondence to:
Jennifer Schaffner
Program Officer
jennifer_schaffner@oclc.org

Suggested citation:
Schaffner, Jennifer, Francine Snyder, and Shannon Supple. 2011. "Scan and Deliver: Managing User-initiated Digitization in Special Collections and Archives." Dublin, OH: OCLC Research. http://www.oclc.org/research/publications/library/2011/2011-05.pdf.

# Acknowledgements

The RLG Partnership Working Group on Streamlining Photography and Scanning created this report collaboratively. We gratefully acknowledge the contributions of each of the following members:

Anne Blecksmith
Getty Research Institute

Eleanor Brown
Cornell University

Paul Constantine
University of Washington

Gordon Daines
Brigham Young University

Tiah Edmunson-Morton
Oregon State University

Cristina Favretto
University of Miami

Steven K. Galbraith
Folger Shakespeare Library

Susan Hamson
Columbia University

Rebekah Irwin
Beinecke Library, Yale University

Sue Kunda
Oregon State University

Jennie Levine Knies
University of Maryland

Suzannah Massen
Frick Art Reference Library

Dennis Massie
OCLC Research

Dennis Meissner
Minnesota Historical Society

Elizabeth McAllister
University of Maryland

Lisa Miller
Hoover Institution Library and Archives, Stanford University

Timothy Pyatt
Duke University

Jennifer Schaffner
OCLC Research

Jon Shaw
University of Pennsylvania

Francine Snyder
Solomon R. Guggenheim Museum

Shannon Supple
University of California at Berkeley

Mattie Taormina
Stanford University

Cherry Williams
Lilly Library, Indiana University

# Contents

# I. Introduction

Increased visibility of special collections has increased interest and use. At the same time, changes in technology have led to expectations that reproduction requests will be fulfilled with digitized images. The result of increased discovery has been a deluge of requests for digital copies.

This report presents strategies for institutions to provide efficient and economical delivery of digital copies of materials in special collections. By making access our top priority,[1] librarians and archivists can fulfill user requests in a more expeditious manner. Conservative, and sometimes justifiable, assumptions about copying rare and unique materials often result in time-consuming, overly-cautious procedures.  These labor-intensive processes and outdated policies can be streamlined to fit both the circumstances of requests and institutional resources.

Delivering digitized versions of materials is now a core function in libraries and archives. Interlibrary loan colleagues consider digitization-on-demand to be routine.[2] In special collections, digitization should be like interlibrary loan, a service that's provided to users on request (often at no cost to the user) and for which a budget and infrastructure are in place. Simplifying an institution's user-initiated digitization workflow will become increasingly critical as researcher expectations continue to gravitate to the digital.

To prioritize user-initiated digitization, the RLG Partnership launched their Working Group on Streamlining Photography and Scanning. Members of the group, faced with shared challenges, chose to work together to reduce cumbersome workflow and policy issues arising from digitizing materials for users.[3] This synthesis is intended to make it easier to meet our own high standards and to encourage evolving use of special collections and archives.

## Shared Challenges

The group pondered the changing nature of access to special collections and digital delivery. Challenges that members of the group faced together included feeling overwhelmed by numbers of requests for a few copies ("one-sies and two-sies"), being unsure about how to store, utilize, and deliver more broadly images that have been digitized for a user, facing shrinking resources, and delivering images to users in a timely manner. How can we adjust

procedures to deliver more images to users more quickly? How can we increase the pace and cut costs at the same time?

The group focused on developing flexible user-initiated digitization workflows in reading rooms, lowering the bar for scanning and metadata standards, and encouraging policies that allow use of personal hand-held cameras. The companion report on cameras in reading rooms, *Capture and Release,* was published in February 2010.[4]

The working group sought common strategies for streamlining the process of creating digital images for users, keeping the images locally, or depositing them in consortial and institutional digital repositories. To evaluate current local practice and locate gaps, we shared our institutions' digitization-on-demand workflows.[5] We discovered common bottlenecks: staff time, metadata, funding, standards, and historical practice. We synthesized workflow strategies and developed a tiered approach to digital reproduction policies and procedures. Along the way we discovered that sound methods to fulfill requests for copies unite public services and technical services in one fluid workflow.

## Delivery Is the Goal

User needs and access to collections must drive all digitization of archives and special collections. In the context of reduced resources and shifting user expectations of online access, a quick and easy way to deliver requested digital reproductions has become an imperative. User requests must not be bogged down by fine-tuning images and metadata. Steps to build digitized collections should be derived from use, but after user requests are fulfilled promptly.

Leveraging user requests is a very effective way of selecting collections to digitize.[6] Mark Custer has called the relationship between selection and user-initiated digitization of special collections "mass representation."[7] His research shows that putting the top 10% of requested images online provides better access for over 50% of users. With a modest amount of additional effort, an institution that places the top 20% of requested images online reaches over 70% of its users. Zipf's Law – the principle of least effort – means that after a point, digitizing more collections has diminishing impact for the time and effort spent.

Efficient and economical user-initiated digitization is not the same in all situations. Different collections, institutional policies, and resource levels inform decisions about user requests.[8] Allowing users to take photographs with their own hand-held cameras requires the least institutional resources. Quickly scanning and delivering a digital copy requires more resources, but not much more. Sometimes copies made for users are worth keeping, sometimes not. In some cases, it is better to digitize a whole folder or volume while an item is in hand. Best

practice standards for digital libraries are always available to institutions that choose to commit to them. When large-scale digitization of special collections is impractical, digitized images originally made for users provide "mass representation" of our best collections. In the future, we can return to the originals and take a different digitization track.

It's all about delivery.

## II. Tracks for User-initiated Digitization

The four main steps in a user-initiated digitization process are embedded in Figure 1, "Tiered Workflow for User-initiated Digitization": review, decide, scan, and deliver. Tracks for digitization workflows can be chosen and combined based on decisions about quality, quantities, standards, metadata, and allocation of resources. Knowing which questions to ask can go a long way to overcoming stumbling blocks and streamlining the process.

A flexible approach acknowledges differences in user needs, collections, institutions, and resources. As always, institutions will use professional judgment to scale up requirements as needed and when they can. Delivery of digitized materials to users is the goal, no matter what combinations of tracks you choose. However, the RLG Working Group firmly recommends taking the inside track first and scaling up when required.

| | INSIDE TRACK | MIDDLE TRACK | OUTSIDE TRACK |
|---|---|---|---|
| **REVIEW** | | | |
| **REQUEST** | Verbal request | Simple form | Detailed forms |
| **APPROVE** | One staff member | Two staff | Three or more staff |
| **SEARCH FOR EXISTING SURROGATE** | Don't take the time | Quick check (local files or online) | Search:<br>• Google Books<br>• EEBO & ECCO<br>• HathiTrust<br>• Etc… |
| **RIGHTS** | User's responsibility | User's responsibility | Search, control, & monitor thoroughly |
| **DECIDE** | | | |
| **STAFF** | One staff member | Two staff | Multiple specialists:<br>• Curator<br>• Conservator<br>• Metadata creator<br>• Image professional<br>• Etc… |
| **RESOLUTION** | Easiest option | Middling to high quality | High quality |
| **METADATA** | None | Filename and keyword or two | Metadata & citation (choose standards) |
| **WHOLE / PART / ITEM** | Will scan whole unit if it takes up to "x" minutes | Entire chapter or series | Entire volume or collection (large-scale digitization) |
| **SCAN** | | | |
| **STAFF** | One staff member | Two staff | Multiple staff |
| **QUALITY CONTROL** | Trust your staff! | Quick review | Thorough review by multiple specialists (color-correction, etc.) |
| **DELIVER** | | | |
| **DELIVERY METHOD** | Deliver forthwith to user | Keep in local files and Web site | • Digital library<br>• Digital repository<br>• DAMS<br>• Consortia<br>• Aggregations |

**Figure 1.  Tiered Workflow for User-initiated Digitization**

After considerable discussion, consensus of the working group is that funding is the biggest factor affecting digitization-on-demand workflows. Other factors that influence decisions are time, staff, users, and collections. A variety of tracks – defined by your objectives and funding – can serve as guiding principles to establish consistent, compact workflows for user-initiated digitization. At each decision, jump from track to track and iterate as needed.

- **Taking the inside track** is often based on stretching the institution's resources. Decisions are made to maximize resources available for user-initiated digitization by using solid baseline practices. The primary focus on the inside track is to get the job done quickly and to fill the user's request.

- **Taking the middle track** has the widest range of options, standards, and results. This is the most flexible of the tracks, where decisions often fall in gray areas.

- **Taking the outside track** focuses on the collections themselves. While users may initiate digitization, it is undertaken to deliver materials to a greater public. These decisions may lead to comprehensive digitization, such as an entire book, series, or collection. The goal is to create maximum access to special collections, using preservation and archival standards. This track usually involves a level of thought and planning that is more in-depth than the fulfillment of day-to-day digitization requests.

Institutions should adopt a flexible approach, jumping tracks within workflows as desired. This allows for an elastic workflow that can be manipulated for a variety of projects and requests. Scan and deliver!

# III. Tiered Workflow

Steps in user-initiated digitization – review, decide, scan, and deliver – have a spectrum of tracks for workflows. The following offer ways to work through the questions involved in making these tiered workflow decisions. In Figure 1, names of your staff can be filled in along a continuum of the inside track, middle track, and outside track.

## A. Review

Review steps may be well-established from photocopying processes, but it is worthwhile to revisit them to look for opportunities to improve efficiency.

1. *How do users request digital reproductions?* Taking the inside track may be as simple as a verbal request. The middle and outside tracks have more steps, ranging from a basic form that includes contact information and copyright acknowledgment, to detailed forms that include a list of items with required metadata. One way to move toward the inside track is to condense requests into one generic form.

2. *How are requests for copies approved?* Institutions on the inside track trust that if the institution owns an item, it warrants digitization.[9]

    a. *How many people need to approve materials to be digitized?* Minimize the number of people involved. Assign one person to decide when to approve simple requests and when to bring in other people. Train your staff and trust their judgment. Depending on resources and the nature of the request, approval queues vary from review by a single staff member to review by public services staff, curators, and conservators. On the inside track, use blanket guidelines to address concerns about the physical condition of originals, such as a policy that fragile or oversize items may not be digitized. On the outside track, a staff conservator might prepare and treat fragile and oversize items before they are scanned.

    > *At Oregon State University, every request goes directly to the digital library. There is one workflow, so no one needs to agonize over decisions.*

    b. *Have the requested items already been digitized?* One of the first steps could be to see if the items have already been scanned.[10] When you deliver existing copies, the request is complete. On the other hand, searching can be very time consuming and the quality may not be the kind needed by the user.

> *Duke University has found that the middle track doesn't make sense for them, so they skip it. They either opt for the inside track, to deliver images to users immediately, or take the outside track. It isn't worth their time to go hunting for local copies.*

c. *How do you manage copyright, privacy, and other legal issues?* The inside and middle tracks place responsibility for rights and privacy issues on the user. The outside track often involves searching to determine the legal status of materials. Some institutions include restrictions in their catalog records; sometimes this is required for digital libraries. Institutions choosing an intermediate track might provide some guidance with a signed agreement. Choose a reasonable approach when digitizing whole collections.[11]

## B. Decide

Once a digitization request is approved, scanning and metadata specifications must be determined.

1. *Will you keep the images?* Whether or not an institution will keep copies of the images makes an enormous difference in the time and resources committed to user-initiated digitization. Choices about further expenditures of time and money must align sensibly with both user needs and your objectives.

a. *Is it useful for you to keep the digitized images?* If you do not have resources to tinker with metadata, do not keep digitized images. Instead, allow users to capture images themselves with hand-held cameras. On the inside track, scan the materials at the minimum quality requested and deliver the copies to the user. These are the most streamlined options for access and can often be the most appropriate.

> *Retire the stack of CDs filled with one-off images that have opaque file names.*

b. *Do you have sufficient resources to keep the digitized images?* Digital infrastructure and metadata creation always require significant resources. On the inside track, do not keep the images. On an intermediate track, digitized images are kept locally (such as on a network drive) for future requests. On the outside track, digitized images are integrated into institutional digital asset management systems (DAMS), consortial repositories, or the institution's digital library. On the middle or outside track, institutions may want to reuse images for projects, such as Web exhibits, the Flickr Commons, or links in online finding aids and catalog records.

> *Stanford University developed a tiered strategy to expand access to special collections. Because some users like the serendipity of browsing through images, they take a middle track to deliver requested images on their local Web site. When users need more, Stanford jumps to the outside track.*

2. *What resolution is needed for scanning?* User needs determine the quality of scans. If you have options and decide to keep copies, though, you might choose a higher quality than what the user requested to meet your own purposes.

a. *Even if the user requests images only for research or personal use*, there are a variety of tracks from which to choose. If you are not retaining copies, rely on your institution's camera policy and allow the user to take their own photographs. To digitize for local reuse, your own specifications are the default. On the outside track, digitizing with publication or archival quality and wrapping images in detailed metadata commits resources to repurpose requested images for network-level access.

> *Cornell University developed a full-blown workflow on the outside track to enlarge their digital library by including digitized images requested by users. But there are exceptions; Cornell doesn't choose the outside track in every case.*

b. *Is the request for high-resolution?* Institutions on the middle track might scan at a high resolution but deliver images with minimal metadata. Sometimes an institution might jump to the outside track and scale up.

3. *What metadata will you create?* If you want to keep the images, metadata is the biggest hurdle in repurposing scans. The amount and type of metadata varies substantially, from a robust schema following national or international standards to a bare-bones schema with minimal metadata, or even just a semantic file name. One economical tactic is to reuse whatever metadata the user included in the request.

> *"Standards are great until they get in your way. I have ignored both internal and external standards, but it gets effective work done."*

a. On the outside track, you face a glut of current digital library standards.[12] You will need to determine what standards you will use and whether you will use persistent identifiers. New metadata can be created and existing metadata can be pulled from the library catalog or finding aid.

> *At the Guggenheim Museum Archives the bulk of scans are created for access, not preservation or publication. They add the minimal metadata needed to drop the images into their finding aids.*

b. Moderate approaches take the middle track between the extremes of using standards extensively and avoiding metadata altogether. Is there existing metadata that can be reused? The minimal metadata required for local access and reuse is nothing more than a few keywords, or the collection number or name.

> *The absolute minimum metadata requirement is no additional metadata at all. Link the digital object (whether simple or complex) to its description in some other resource, like a catalog record or finding aid. This way, it inherits the metadata of the original, in context.*

4. ***Will you scan the whole volume, folder, or series when the request is for an item or set of items?*** Questions to ask when choosing which track to take are: Is it cost-effective, are you outsourcing, does the part make little sense without the whole, is there great research interest, and are you keeping images?

   a. Sometimes it doesn't make sense to digitize just one item or page. On the inside track, however, often one image will do. You may also opt for a policy that says you scan the whole unit if it takes less than a given amount of time (e.g., 15 minutes).

   b. If taking the middle or outside track, scan the whole unit. On the middle track, this might be the entire folder or chapter. If you're investing a lot of time and resources in digitization, the entire series or collection might have high research value or see high use.

## C. Scan

Using a scanner may be fast and easy, but managing digitization requires professional judgment. Deciding ahead of time about how much to invest in requests for reproductions can stabilize management of metadata and digitization.

1. ***Who digitizes and describes materials and how many steps are involved?*** Streamline digitization workflow so that one or two people are involved in the whole process, not six or seven.

   a. Taking the outside track may require many staff, including curators, public services staff, technology staff, conservators, metadata specialists, etc. However, fewer staff may be required, even on the outside track. The middle track calls for a smaller group of staff, while the inside track may involve only

one staff member. In some situations, the same person reviews the request and scans the items.

   b.  After you have reduced the number of staff involved, it helps to look at digitization workflow critically and evaluate its efficiency. For example, how close together are storage spaces, reading rooms, and scanners? How are materials transported from one point to another?

2.  ***Will you perform quality control on the images, metadata, or originals?*** Quality control is often viewed as a tight bottleneck in digitization workflows.

   a.  The outside track usually includes quality control and assessments. The middle track may include some review for quality, such as quick checks of images and metadata. The inside track achieves the greatest efficiency because staff do not need to control images or metadata at all.

   b.  The outside track often incorporates post-digitization condition assessments for the originals. On the other hand, on the inside track you trust your staff and your existing policies for handling materials and making copies.

## D. Deliver

1.  *How do you deliver digital copies to the user?*

   a.  Delivery on the outside track is most often automated and online. Complex choices for delivery vary, from a Web-based solution such as Flickr, to an enterprise system like CONTENTdm, or to a locally-managed open-source repository like FEDORA. Use-derivations with descriptive metadata, created according to institutional specifications and best practices, are published to the institution's DAMS or preservation repository. In some cases users can find these images on their own and need not make requests. We strongly recommend that digitized images be linked to both their finding aids and catalog records, and whatever alternative discovery sources your institution provides.

   b.  On the middle track, delivery could be partially automated, but might require some assistance. For example, some institutions store digitized images in local files or deliver them via the Web. Some institutions have an online gallery populated with previous users' requests.

c. To deliver digitized images to users quickly, the inside track involves emailing them or copying them to CDs, FTP sites, or flash drives. That's it.

> *Cornell University loads digitized images of special collections to Flickr Commons. Duke University puts images into iTunes. The University of Pennsylvania partners with Kirtas to deliver high-quality images via print-on-demand. Stanford University gives digitized images of special collections to Google. Workflows on the outside track fit when the extra cost and effort meet user needs.*

# IV. Conclusion

The RLG Partnership Working Group on Streamlining Photography and Scanning believes that principles from low-overhead and moderate workflows for user requests can be incorporated into large-scale digitization programs. In special collections and archives, we can choose flexible tracks to increase use of digitized collections. Resource allocators must fund user-initiated digitization directly and allow use to inform digital collection building.

Workflows appropriate to individual institutions need not be paralyzing when managed in flexible and economical ways. Prevent overburdened gridlock by streamlining user-initiated digitization workflows to the most appropriate track(s). Embrace the opportunity to leverage digital technology to deliver special collections and archives to users efficiently.

Scan and deliver.

# Appendix A: A Selection of Digital Library Standards compiled by Jennie Levine Knies

For ideas about best practice, see:

Bibliographic Center for Research. CDP Digital Imaging Best Practices Working Group.  2008. *BCR's CDP digital imaging best practices*, *version 2.0*. Aurora, CO: Bibliographic Center for Research.

*Descriptive metadata guideline for RLG Cultural Materials.* 2005. Mountain View, CA: RLG. http://www.oclc.org/research/activities/past/rlg/culturalmaterials/RLG_desc_metadata.pdf.

Federal Agencies Digitization Initiative (FADGI) – Still Image Working Group. 2010. Rev. ed. *Technical guidelines for digitizing cultural heritage materials: creation of raster images and master files.* http://www.digitizationguidelines.gov/guidelines/FADGI_Still_ImageTech_Guidelines_2010-08-24.pdf

Gueguen, Gretchen. 2007. *Digital imaging primer*. College Park, MD: University of Maryland Libraries. http://www.lib.umd.edu/dcr/publications/DigitalImagingPrimer.pdf

Library of Congress, American Memory. N.d. *Building digital collections: a technical overview.* http://memory.loc.gov/ammem/about/techIn.html

Luong, Giao. 2009.  *Digital imaging overview.* Los Angeles, CA:  University of Southern California Digital Library http://www.usc.edu/libraries/collections/digitallibrary/documents/USCDL_Digital_Imaging_Overview.pdf

*Moving theory into practice: digital imaging tutorial.* 2003. Ithaca, NY: Cornell University. http://www.library.cornell.edu/preservation/tutorial/

National Information Standards Organization (NISO) Framework Working Group. 2007. 3rd ed. *A Framework of guidance for building good digital collections.* http://framework.niso.org/

New York State Archives. 2006. *Imaging production guideline*s. http://www.archives.nysed.gov/a/records/mr_erecords_imgguides.shtml

Petersen, Kit A. 2004. *Standards related to digital imaging of pictorial materials.* Washington DC:  Library of Congress. http://www.loc.gov/rr/print/tp/DigitizationStandardsPictorial.pdf

# Notes

[1] *Digitization of special collections: RBMS principles for digital content*,
http://rbms.info/committees/task_force/digitization/digprinciples.html. *ACRL/SAA joint statement on access to research materials in archives and special collections libraries*,
http://www.ala.org/ala/mgrps/divs/acrl/standards/jointstatement.cfm and
http://www.archivists.org/statements/ALA-SAA-Access09.asp

[2] Unpublished results from a workflow survey conducted for the project *Sharing special collections*,
http://www.oclc.org/research/activities/sharing/default.htm. Reference to preliminary results of the survey are at http://www.oclc.org/research/events/2010-06-09a.htm#sc.

[3] The project description is online at: http://www.oclc.org/research/activities/photoscan/default.htm.

[4] Lisa Miller, Steven K. Galbraith, et al. 2010. *"Capture and release": digital cameras in the reading room.* Dublin, OH: OCLC Research. http://www.oclc.org/research/publications/library/2010/2010-05.pdf.

[5] We analyzed descriptions of digitization-on-demand workflows from: University of Maryland, College Park; Hoover Institution Library and Archives, Stanford University; Duke University; L. Tom Perry Special Collections, Brigham Young University; Oregon State University; Folger Shakespeare Library; Lilly Library, Indiana University; University of Washington; Solomon R. Guggenheim Museum Archives; Getty Research Institute; Cornell University; University of Miami; National Archives of Australia; Clark Library, UCLA.

[6] In 2007, Nick Poole argued that "digitisation-on-demand" is economical and scalable, while "The economics of mass-digitisation are inherently unsustainable for cultural organisations." Presentation available online at:
http://www.den.nl/docs/20071203103043/#Keynote2:NickPoole%28MDA%29%22WhatAudience?Thedeathofmass-digitisationandtheriseofthemarketeconomy%22. See also a summary of the entire keynote panel at the Digitaal Erfgoed Conferentie in *Informatie professional*, 2008 (February): 12-13.

[7] Mark Custer. 2009. "Incorporating patron requests into archival workflows and digital repository interfaces." Presentation available online at:
http://saa.archivists.org/Scripts/4Disapi.dll/4DCGI/events/eventdetail.html?Action=Events_Detail&InvID_W=1089. Forthcoming as an article with the title, "Mass representation defined: a study of patron requests."

[8] Decisions outside the scope of this document – regarding fees charged to patrons, limitations on requests, intellectual property issues, and donor guidelines – follow individual institutional policies.

[9] Ricky Erway and Jennifer Schaffner. 2007. *Shifting gears: gearing up to get into the flow.* Dublin, OH: OCLC Research. http://www.oclc.org/programs/publications/reports/2007-02.pdf.

[10] Self-evident choices are, of course, Google books, the Open Content Alliance, an institution's local repository, consortia, vendor products, etc.

[11] *Well-intentioned practice for putting digitized collections of unpublished materials online.* 2010.
http://www.oclc.org/research/activities/rights/practice.pdf.

[12] See Appendix A.